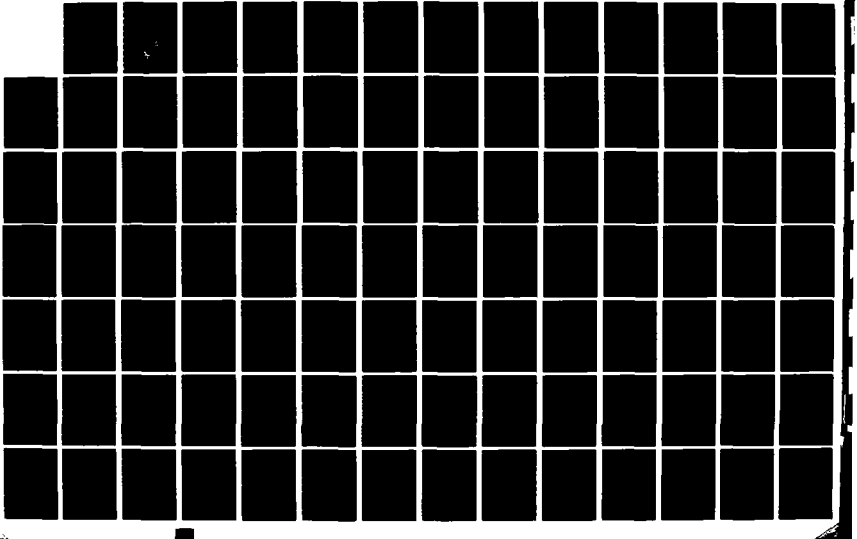
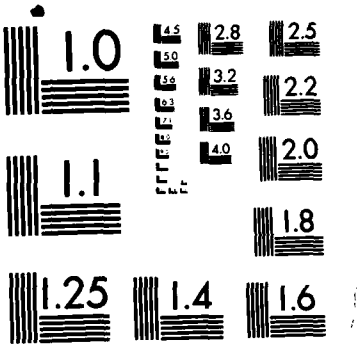
AD-A162 225 PERFORMANCE ANALYSIS FOR HYBRID STATE ESTIMATION PROBLEMS(U) ALPHATECH INC BURLINGTON MA R B WASHBURN ET AL OCT 85 N00014-84-C-0370 1/2

UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

(12)

TR-266

PERFORMANCE ANALYSIS FOR HYBRID STATE ESTIMATION PROBLEMS

FINAL REPORT

ALPHATECH, INC.
2 BURLINGTON EXECUTIVE CENTER
111 MIDDLESEX TURNPIKE
BURLINGTON, MA 01803
617-273-3388

ALPHATECH, INC.

TR-266

PERFORMANCE ANALYSIS FOR HYBRID STATE ESTIMATION PROBLEMS

FINAL REPORT

By

R.B. Washburn
T.G. Allen
D. Teneketzis

October 1985

Submitted to:

Mr. R.J. Simpson
Code 411 MA
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

Contract No. N00014-84-C-03708

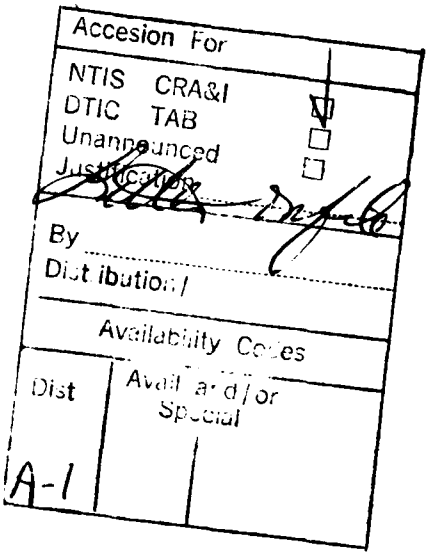
ALPHATECH, Inc.
2 Burlington Executive Center
111 Middlesex Turnpike
Burlington, Massachusetts 01803
(617) 273-3388

ABSTRACT

Hybrid state estimation problems are statistical estimation problems in which both continuous-valued and discrete-valued states and parameters occur. The hybrid state model provides both a natural formulation for many types of surveillance and tracking problems and a powerful framework for deriving theoretically optimal and practical suboptimal tracking algorithms. This report describes research in developing tools for evaluating the performance of optimal hybrid state estimation problems.

ACKNOWLEDGEMENT

This research was supported by Code 340R of the Naval Air Systems Command under contract N00014-84-C-03078 with the Office of Naval Research.



CONTENTS

CONTENTS (Continued)

SECTION 1

INTRODUCTION

Although the mathematical structure of the optimal solution of the multiobject tracking problem is well understood, the excessive computational complexity of the optimal tracking algorithm prohibits its practical realization using even the largest and fastest computers available in the foreseeable future. For this reason, many different tracking algorithms have been developed which sacrifice optimal performance for the sake of computational feasibility. Intuition suggests that some suboptimal approximations to the optimal algorithm result in little performance loss, but there exists no reliable quantitative analysis of suboptimal performance that shows this is, in fact, the case.

To a limited extent, it is possible to compare one suboptimal algorithm to another suboptimal algorithm by numerically simulating the performance of both. Such simulations provide useful information about the relative performance of different suboptimal algorithms, but they do not give reliable quantitative measures of performance without extensive simulation. In all but the simplest multiobject tracking scenarios, it is expensive and time-consuming to compute an adequate number of simulation samples for reliable Monte-Carlo analysis. For this reason, it is difficult to use simulation to investigate how an algorithm's performance depends on algorithmic and environmental parameters.

While simulating relative performance of suboptimal algorithms is difficult, simulating optimal performance to determine the absolute performance of a suboptimal algorithm is effectively impossible. In all but the simplest cases, it is computationally impossible to compute even one simulation sample of the optimal tracking algorithm, let alone an adequate number of samples for Monte-Carlo analysis. Although one can obtain some insight by comparing simulated performance of optimal and suboptimal algorithms for a small number of very limited cases, it is not possible to obtain quantitative measures of performance in this way.

What are lacking are analytical methods for analyzing the quantitative performance of optimal and suboptimal multiobject tracking algorithms. Because of the excessive computational requirements of simulating such algorithms, there is no feasible alternative to analytical methods of performance analysis. Such analytical methods do exist and have proven useful for analyzing performance in other, much simpler types of hypothesis testing and estimation problems. There, efficient analytical performance analysis techniques have greatly facilitated the design of practical algorithms. The basic research problem here is to find similarly effective analytical techniques for the more difficult hypothesis testing and estimation problems of multiobject tracking.

We base our approach to this problem on studying multiobject tracking problems as a special case of the general class of hybrid state estimation problems [1],[2]*. Hybrid state estimation problems are statistical estimation problems in which both continuous-valued and discrete-valued states and parameters occur. Thus, hybrid state estimation includes both classical

*Reference are indicated by numbers in square brackets. The list appears at the end of the main body of this report.

parameter estimation and hypothesis testing as special cases. In our previous work [1],[2], we have found that the hybrid state model provides both a natural formulation for all types of multiobject tracking problems and a powerful framework for deriving optimal and suboptimal tracking algorithms. The importance of the hybrid state system viewpoint in the research described here is that it gives us a bridge between the simpler estimation and hypothesis testing problems for which effective performance analysis techniques exist and the more difficult estimation and hypothesis testing problems of multiobject tracking for which effective techniques have yet to be developed.

Our overall approach is to consider a hierarchy of hybrid state estimation problems of increasing difficulty, starting at relatively simple problems whose solutions are readily obtained and leading to more difficult multiobject tracking problems. To carry this out, it is convenient to study three different types of hybrid state estimation problems. In the first type (Type I) of problem we are interested only in estimating continuous-valued states and discrete-valued states enter the problem only as measurement noise. Type I problems include those in which the number of targets is known but the association of measurements with targets and false alarms is not known. Since Type I problems are continuous state estimation problems, one can try to extend many different available methods of performance analysis to this type of problem. What is unconventional about such problems (and tends to make performance analysis difficult) is that the discrete-valued noise is non-Gaussian (being a point process in nature) and nonadditive (often being multiplicative).

The second type (Type II) of hybrid state estimation problem generalizes the Type I problem in the following way. In this problem we are still interested in estimating continuous states, but the discrete states enter into the

problem in a much more fundamental way than in Type I problems. In particular, the discrete states are assumed to form a Markov process. Note that we assume that the continuous states cannot affect the discrete states, although the discrete states can affect the continuous states. Problems of tracking targets which can maneuver, or estimating parameters in a system whose dynamics change abruptly (e.g., due to failures) can be formulated as Type II hybrid state estimation problems. Finally, we identify a third type of hybrid state estimation problem (Type III) which are the same as Type II problems except that the estimation problem includes the estimation of discrete states as well as continuous states. Type III problems include the most realistic and most difficult multiobject tracking problems in which the number of targets is unknown and one is interested in both tracking and detecting targets entering and leaving a surveillance area.

This report describes the results of our research in analyzing the performance of hybrid state estimation problems. Section 2 describes the general hybrid state model of interest and several concrete examples relevant to tracking problems; it also defines a classification (Types I, II, III) of hybrid state estimation problems. Section 3 analyzes a simple hybrid state estimation problem in detail - indicating the difficulties inherent in determining optimal performance for this class of problems. Section 4 describes a Cramer-Rao lower bound on mean square error for a large class of Type I hybrid state problems. In Section 5 we discuss a rate distortion approach for Type II problems, and in Section 6 we describe a Monte-Carlo approach based on the representation theorem which applies to very general hybrid state problems. The results of Sections 3 through 6 apply to finite dimensional hybrid state

problems described in Section 2. Extension to hybrid state problems of multiobject tracking is the subject of Section 7. There we introduce a random point process model of multiobject tracking and discuss its application to tracking and performance analysis. Section 8 concludes the report and discusses some further research problems.

SECTION 2

HYBRID STATE ESTIMATION

2.1 GENERAL HYBRID STATE MODEL

Multiobject tracking problems involve both the estimation of continuous-valued parameters such as target positions and velocities, and the the testing of discrete hypotheses such as the association of measurement data with targets. Thus, it is natural to pose the problem of multiobject tracking as a hybrid state estimation problem -- that is, estimation for a partially observed Markov process with discrete- and continuous-valued states. In our previous work [1], [2] we have shown that the hybrid state viewpoint provides a powerful framework in which to formulate all types of multiobject tracking problems. The research reported here is continuing this approach by developing performance analysis techniques for several types of hybrid state estimation problems relevant to multiobject tracking.

To focus our research, we will consider a particular class of hybrid state system, which we call Gauss-Markov hybrid state systems. This is the class of systems which can be described by the equations

$$x(t+1) = A(q(t))x(t) + B(q(t))w(t) \qquad (2\text{-}1a)$$

$$y(t) = C(q(t))x(t) + D(q(t))v(t) \qquad (2\text{-}1b)$$

where the variables are defined so that

$q(t)$ = finite state Markov chain with time invariant transition probabilities;

$x(t)$ = n-dimensional state process;

$w(t)$ = p-dimensional white Gaussian state process noise;

$y(t)$ = m-dimensional measurement process;

$v(t)$ = r-dimensional white Gaussian measurement noise.

For a fixed value of q of the Markov chain, $A(q)$, $B(q)$, $C(q)$, and $D(q)$ are matrices with dimensions corresponding to the vectors described above. One important point to note about the model described by Eq. 2-1 is that the discrete state process $q(t)$ may drive the continuous state $x(t)$ but not conversely. This assumption simplifies analysis and still permits the treatment of many (but not all, see [2]) realistic multiobject tracking problems.

We have assumed that the state and measurement equations are linear in the continuous state variable x in order to focus on the essential hybrid state aspects of our problem. Some of the methods we discuss will extend to nonlinear hybrid state problems of the general form

$$x(t+1) = f\big(x(t),q(t)\big) + B\big(q(t)\big)w(t) \qquad (2\text{-}2a)$$

$$y(t) = h\big(x(t),q(t)\big) + D\big(q(t)\big)v(t) \qquad (2\text{-}2b)$$

where $B(q)$ and $D(q)$ are matrices as above, but $f(x,q)$ and $h(x,q)$ are now allowed to be nonlinear functions of x for each fixed value of q. The nonlinear model of Eq. 2-2 allows the treatment of range-azimuth measurement nonlinearities and passive tracking measurement nonlinearities, but the resulting performance analysis problem will include all the usual difficulties of analyzing nonlinear estimation problems in addition to the difficulties of

analyzing hybrid state problems. In the next subsection we consider several examples of Gauss-Markov hybrid state models relevant to tracking.

2.2 TRACKING EXAMPLES OF HYBRID STATE MODELS

We will give four examples in this subsection to illustrate how one may use the Gauss-Markov hybrid state equations of 2-1 to model tracking problems. These models are simplifications of real tracking problems; in particular, they approximate complex transducers and signal processors with simple hybrid state statistical models. We have chosen this type of approximation to focus our attention on the "backend" tracking problem (which is the object of our research) and to avoid less important details of "frontend" processing.

The first example is a simple model of a sensor and a single target. The sensor produces one measurement per time period; each measurement is either a measurement of the target's state together with background noise, a measurement of independent random clutter, or a measurement of ambient background noise alone. The target moves according to a simple linear Gaussian stochastic difference equation. The hybrid state model for this example is given as follows.

$$x_1(t+1) = x_1(t) + x_2(t) \tag{2-3a}$$

$$x_2(t+1) = x_2(t) + w(t) \tag{2-3b}$$

$$y(t) = q_1(t)[x_1(t) + v_1(t)] + [1-q_1(t)]q_2(t)v_2(t) \tag{2-4}$$

Eq. 2-3 is the state equation and is the usual random acceleration model in which x_1 denotes the target position and x_2 denotes the target velocity. The white noise w is thus an assumed random acceleration. Eq. 2-4 is the measurement equation for this example and contains all the discrete state dependences.

We assume that $q_1(t)$ and $q_2(t)$ are independent discrete state random variables taking variables taking the values 0 or 1. Assume that $q_1(t)$ equals 1 with probability p_i. Then for each time period t, the measurement y(t) is either

a target measurement $x_1(t)$ with some error $v_1(t)$ with probability p_1;

a clutter measurement $v_2(t)$ with probability $(1-p_1)p_2$;

or the 0 measurement with probability $(1-p_1)(1-p_2)$.

Thus, p_1 is the probability of detecting the target, $(1-p_1)p_2$ is the probability of detecting a false alarm, and $(1-p_1)(1-p_2)$ is the probability of detecting nothing at all. These probabilities might be selected to correspond to the receiver operating characteristics of a particular sensor and signal processor frontend. The problem of this example is to estimate $x_1(t)$ and $x_2(t)$, the target position and velocity states. As we will see below, the discrete variables $q_1(t)$ and $q_2(t)$ are not desired and do not have to be estimated in this example although a practical tracking algorithm might in fact have a detector that would estimate these discrete variables. This point will be discussed in the next subsection where we classify hybrid state estimation problems more precisely into three types (Types I, II, and III).

The second example is a variation of the first to include data association ambiguities in the measurement model. In this example there are two targets with states x_1,x_2 and x_3,x_4, respectively. The sensor observes two returns y_1 and y_2, but it does not observe the origin of the measurements. The Gauss-Markov hybrid state model is given by

$$x_1(t+1) = x_1(t) + x_2(t) \qquad (2\text{-}5a)$$

$$x_2(t+1) = x_2(t) + w_1(t) \qquad (2\text{-}5b)$$

$$x_3(t+1) = x_3(t) + x_4(t) \tag{2-5c}$$

$$x_4(t+1) = x_4(t) + w_2(t) \quad . \tag{2-5d}$$

$$y_1(t) = q(t)x_1(t) + \big(1-q(t)\big)x_2(t) + v_1(t) \tag{2-6a}$$

$$Y_2(t) = \big(1-q(t)\big)x_1(t) + q(t)x_2(t) + v_2(t) \tag{2-6b}$$

Equation 2-5 is the usual state model for two independent targets. In Eq. 2-6 we assume that q(t) takes the values 0 or 1, both with probability 1/2. Thus, the sensor measures both target positions, but the ordering of these measurements is randomly permuted.

The third example is an extension of the first to allow the target to make significant maneuvers at unpredictable times. The measurement model Eq. 2-4 remains the same ,but the state model becomes the hybrid state equation

$$x_1(t+1) = x_1(t) + x_2(t) \tag{2-7a}$$

$$x_2(t+1) = x_2(t) + q_3(t)w(t) \tag{2-7b}$$

In Eq. 2-7b the discrete state process $q_3(t)$ is a Markov chain taking two values Q_1 and Q_2 with transition probabilities P_{ij} of jumping from discrete state Q_i to state Q_j. The new discrete state models two types of random acceleration distinguished by different covariances. For example, the covariance Q_1 might have a small value representing small perturbations of essentially constant velocity motion and the covariance Q_2 might have a large value representing occasional significant target maneuvers. By appropriately selecting the transition probabilities P_{ij}, one can adjust the average time between maneuvers. Note that the problem in this example is to estimate $x_1(t)$ and $x_2(t)$. It may be necessary to estimate $q_3(t)$ also, but that is only

incidental to the main tracking objective. In a sense, $q_3(t)$ is a nuisance parameter as far as tracking is concerned. Note that in a different version of this problem, $q_3(t)$ might also be desired (e.g., this would be the case if detecting maneuvers implied significant tactical information such as that a missile launch was about to occur).

The final example is another extension of the first example in which we now model the appearance and disappearance of the single target. Let $q_3(t)$ denote a Markov chain taking two values 0 and 1 with transition probabilities P_{ij} of jumping from discrete state i to state j. When $q_3(t) = 1$, we consider that the target is present and when $q_3(t) = 0$, we consider that the target is absent. The state equation for this example is the same as Eq. 2-3; the measurement equation becomes the following.

$$y(t) = q_1(t)[q_3(t)x_1(t) + v_1(t)] + (1-q_1(t))q_2(t)v_2(t) \qquad (2\text{-}8)$$

The problem in this example, unlike the preceding two examples, is to estimate $q_3(t)$, as well as $x_1(t)$ and $x_2(t)$. That is, it is desired to know whether the target is present (i.e., observable) as well as what are its position and velocity. Indeed, in some cases the position and velocity may only be relevant if $q_3(t) = 1$.

2.3 CLASSIFICATION OF HYBRID STATE ESTIMATION PROBLEMS

It is convenient to define three types of hybrid state estimation problems based partly on the structure of the hybrid state model and partly on the desired estimation criterion. Type I problems are modeled by hybrid state equations of the form of Eq. 2-1 where the discrete state process q(t) is a sequence of independent finite state random variables. The estimation problem

is to estimate the state x(t) given measurements y(s) up to time t. The important point to note is that q(t) plays the role of a nuisance parameter or an additional noise process analogous to w(t) and v(t). It is not necessary to estimate q(t) any more than it is necessary to estimate w(t) or v(t) to obtain an estimate of x(t). This is not to say that some estimation algorithm may, in fact, estimate q(t) as an intermediate step in obtaining an estimate of x(t).

The significance of Type I problem lies in the fact that they are essentially continuous state estimation problems and therefore, many classical performance analysis techniques (such as Cramer-Rao type methods) apply to this type of problem, at least in theory. As we will see later in this report, there is a considerable difficulty in applying the theory because the distributions are non-Gaussian (being a mixture of Gaussian distributions).

We will often consider a subclass of Type I problems for which the continuous state dynamics are independent of the discrete variables, namely Eq. 2-1 is replaced by

$$x(t+1) = Ax(t) + Bw(t) \tag{2-9a}$$

$$y(t) = C\big(q(t)\big)x(t) + D\big(q(t)\big)v(t) \tag{2-9b}$$

Even within this subclass of Type I problems it is possible to consider many different tracking examples of interest. The first and second examples of the previous subsection are members of this class of Type I problem. One can easily generalize these examples to model problems with multiple targets and multiple returns per time period, provided that the number of targets is constant and known (i.e., detecting a target's presence or absence is not a part of the problem).

Type II problems are given by the same equations (namely Eq. 2-1), but the discrete state process q(t) is allowed to be a finite state Markov chain rather than a sequence of independent finite state random variables. Nevertheless, it is still desired to estimate only the continuous state x(t) in Type II problems, although even more so than in Type I problems will it be necessary to estimate the discrete state q(t) as an intermediate step in obtaining the estimate of x(t). The second example of tracking a maneuvering target is a Type II hybrid state estimation problem.

Let us make the distinction between Type I and II clear. In Type I problems the joint process $\langle x(t),y(t)\rangle$ is Markov. In Type II problems this is no longer true; the Markov process is the joint process $\langle x(t),y(t),q(t)\rangle$ (or at least some component of q(t)). This property of Type II problems complicates performance analysis considerably. However, the estimation criterion depends just on x(t); for example, it is a mean square error criterion for estimating x(t). This assumption at least gives us a simple measure of performance to consider.

A subclass of Type II problems has the property that q(t) is equal to q(0) for all times t. These are the so-called multiple model problems sometimes used in adaptive estimation. We will not study this class of problems and refer to reader to [3] for further information.

Type III problems generalize the Type II problem by allowing the estimation criterion to depend on both x(t) and on q(t). This type of hybrid state estimation problem also has the model of Eq. 2-1 in which q(t) is allowed to be a general finite state Markov chain. The third example of tracking a single target which appears and disappears at random times is a Type III problem if we are interested in knowing when the target is present and when it is absent.

Type III problems include the most realistic multiobject tracking problems in which the total number of targets is unknown and in which one is interested in some aspect of the identity of the target (if only to know that an object being tracked at the current time is the same as an object that was being tracked at an earlier time; that is, if track continuity is an issue). In such problems it is difficult to define meaningful measures of performance, especially if they are to be amenable to analysis. In the rest of this paper we will discuss performance analysis for different hybrid state estimation problems.

SECTION 3

ANALYSIS OF A SIMPLE PROBLEM

3.1 INTRODUCTION

A simple hybrid state estimation problem of interest is described by the measurement equation

$$y = x + qv_1 + (1-q)v_0 \qquad (3\text{-}1)$$

where x, v_1 and v_0 are normal random variables and $q = 0$ or 1. Assume that x, v_1, v_0 and q are independent and assume tht the distributions are

$$x \sim N(\bar{x}, \sigma^2)$$

$$v_1 \sim N(0, \sigma_1^2)$$

$$v_0 \sim N(0, \sigma_0^2)$$

$$P\{q=1\} = \varepsilon$$

We are interested in approximating the minimum mean square error of estimating x given the measurement y. Equation 3-1 presents an example of estimation with non-Gaussian measurement error. This is true in general for hybrid state estimation problems.

The measurement error

$$qv_1 + (1-q)v_0$$

is a random mixture of two different random variales (v_1 and v_0). For example, this model might represent the problem of estimating x given a probability ε

of taking a bad measurement (with large variance σ_1^2) and probability $1 - \varepsilon$ of taking a good measurement (with small variance σ_0^2).

Suppose $\hat{x}$ is the conditional mean of x given by

$$\hat{x} = E\{x|y\} \quad .$$

Then the minimum mean square error V is given by

$$V = E\{(x-\hat{x})^2\} \quad .$$

Our general goal is to develop good methods for approximating V in hybrid state estimation problems. In this section we discuss several approximations to develop some feeling for different methods' computability and accuracy. In the following subsections we consider direct calculation of V, asymptotic expansion of V (for small ε), the Cramer-Rao lower bound of V, and a rate distortion lower bound of V. We also consider numerical examples of the different approximations to study their relative accuracy.

3.2 ASYMPTOTIC EXPANSION OF OPTIMAL PERFORMANCE

3.2.1 Exact Expression of Minimum Mean Square Error

For the class of Gauss-Markov hybrid state systems defined in Section 2, it is easy to describe the minimum mean square estimator; namely the conditional mean. For this simple problem we have

$$x(y) = E\{x|y, q=0\}\, P\{q=0|y\} + E\{x|y, q=1\}\, P\{q=1|y\}$$

where $E\{x|y, q=0\}$ and $E\{x|y, q=1\}$ are the solution of linear least squares problems, and $P\{q=0|y\}$ and $P\{q=1|y\}$ are easily obtained from Bayes' rule. For the problem at hand

$$E\{x|y,\ q=i\} = \frac{\sigma_i^2\,\bar{x} + \sigma^2 y}{\sigma_i^2 + \sigma^2}$$

$$P\{q=i|y\} = \frac{[2\pi(\sigma^2 + \sigma_i^2)]^{-1/2} \exp\{-1/2\ (y-\bar{x})^2(\sigma^2 + \sigma_i^2)^{-1}\}\ P\{q=i\}}{[2\pi(\sigma^2 + \sigma_1^2)]^{-1/2} \exp\{-1/2\ (y-\bar{x})^2(\sigma^2 + \sigma_0^2)^{-1}\}\ P\{q=1\} + [2\pi(\sigma^2 + \sigma_0^2)]^{-1/2} \exp\{-1/2\ (y-\bar{x})^2 (\sigma^2 + \sigma_0^2)^{-1}\}\ P\{q=0\}}$$

Thus, we can write $\hat{x}$ as

$$\hat{x} = \bar{x} + \xi \frac{(1-\varepsilon)b_0\ \phi_0(\xi) + \varepsilon\ b_1\ \phi_1(\xi)}{(1-\varepsilon)\ \phi_0(\xi) + \varepsilon\ \phi_1(\xi)} \tag{3-2}$$

where

$$\xi = y - \bar{x}$$

$$b_i = \sigma^2(\sigma^2 + \sigma_i^2)^{-1} \tag{3-3}$$

$$\phi_i(\xi) = [2\pi(\sigma^2 + \sigma_i^2)]^{-1/2} \exp\{-1/2\ \xi^2(\sigma^2 + \sigma_i^2)^{-1}\} \tag{3-4}$$

However, we are not interested in $\hat{x}$ itself, but in its mean square performance V. This is found from

$$\begin{aligned} V &= E\{(x-\bar{x})^2\} - E\{(\hat{x}-\bar{x})^2\} \\ &= \sigma^2 - I \end{aligned} \tag{3-5}$$

where $I = E\{(\hat{x}-x)^2\}$ is the integral

$$I = \int_{-\infty}^{\infty} \xi^2 \frac{[(1-\varepsilon)b_0\ \phi_0(\xi) + \varepsilon\ b_1\ \phi_1(\xi)]^2}{(1-\varepsilon)\phi_0(\xi) + \varepsilon\ \phi_1(\xi)}\, d\xi\ . \tag{3-6}$$

Even for this simple problem, the integral I does not have a closed form solution and we must resort to numerical or analytic approxmation of it. This type of integral, in which a sum of Gaussian distribution appears as a denominator of the integrand, is common to all Gauss-Markov hybrid state problems. It occurs in the exact expression of minimum mean square error and in the expression for the Cramer-Rao bound as we will show later. Thus, it is important that we learn to approximate such integrals well. In the next subsection we consider asymptotic expansions of I for small values of ε.

3.2.2 Asymptotic Expansion of Minimum Mean Square Error

By rewriting the numerator in Eq. 3-6 as

$$[(1-\varepsilon)b_0\ \phi_0 + \varepsilon\ b_1\ \phi_1]^2 = [b_0[(1-\varepsilon)\phi_0 + \varepsilon\ \phi_1] + \varepsilon(b_1-b_0)\phi_1]^2 \quad ,$$

we find that

$$\begin{aligned} I = {} & b_0^2 \int \xi^2[(1-\varepsilon)\phi_0(\xi) + \varepsilon\ \phi_1(\xi)]d\xi \\ & + 2\varepsilon\ b_0(b_1-b_0) \int \xi^2\ \phi_1(\xi)d\varepsilon \\ & + \varepsilon^2(b_1-b_0)^2 \int \frac{\xi^2\ \phi_1(\xi)^2}{(1-\varepsilon)\phi_0(\xi) + \varepsilon\ \phi_1(\xi)}\ d\xi \end{aligned} \tag{3-7}$$

The first two integrals in Eq. 3-7 can be computed exactly; the third cannot. Equation 3-5 and 3-7 give us the following asymptotic expansion of the minimum mean square error V.

$$V = \frac{\sigma^2\ \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon\ \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} - \varepsilon^2\ \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]^2}{[\sigma^2 + \sigma_0^2]^2\ [\sigma^2 + \sigma_1^2]^2}\ J \tag{3-8}$$

where

$$J = \int_{-\infty}^{\infty} \frac{\xi^2 \ \phi_1(\xi)^2}{(1-\varepsilon)\phi_0(\xi) + \varepsilon \ \phi_1(\xi)} \, d\xi \quad . \tag{3-9}$$

Since the third term of Eq. 3-8 is negative, we immediately obtain an upper bound of V which we denote by V_1^+:

$$V_1^+ = \frac{\sigma^2 \ \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} \tag{3-10}$$

A lower bound is obtained by noting that

$$J < \frac{1}{\varepsilon} \cdot \int_{-\infty}^{\infty} \xi^2 \ \phi_1(\xi) d\xi = \varepsilon^{-1}[\sigma^2 + \sigma_1^2] \quad . \tag{3-11}$$

Thus, the lower bound is

$$V_1^- = \frac{\sigma^2 \ \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} \left(1 - \frac{\sigma_1 - \sigma_0^2}{\sigma^2 + \sigma_1^2} \right) \quad . \tag{3-12}$$

Let us now study the third term of Eq. 3-8. First note that if

$$\sigma_0^2 < \sigma^2 + 2\sigma_1^2$$

then the integral

$$\int_{-\infty}^{\infty} \xi^2 \frac{\phi_1(\xi)^2}{\phi_0(\xi)} \, d\xi = \infty$$

diverges and

$$\lim_{\varepsilon \to 0} \ J = \infty \quad . \tag{3-13}$$

On the other hand, from Eq. 3-11 it is clear that

$$\lim_{\varepsilon \to 0} \varepsilon J = 0 \quad . \tag{3-14}$$

Equations 3-13 and 3-14 imply that the higher order term $\varepsilon^2 J$ in Eq. 3-8 decreases more slowly than ε^2 but more rapidly than ε. In Appendix A we show that in the case $\sigma \approx \sigma_1 \gg \sigma_0$, the correction term has an asymptotic behavior like

$$\varepsilon^2 J \sim |\ell n\ \varepsilon|^{1/2}\ \varepsilon^{1+\gamma} \tag{3-15}$$

for small ε, where

$$\gamma = \frac{\sigma^2 + \sigma_0^2}{\sigma_1^2 - \sigma_0^2} \quad . \tag{3-16}$$

Appendix A also gives expressions for a second-order upper bound V_2^+ and a second-order lower bound V_2^-. As Appendix A makes evident, detailed analytic approximation of V can be difficult for even very simple hybrid state estimation problems.

3.3 CRAMER-RAO LOWER BOUND

Van Trees' [4] extension of the classical Cramer-Rao bound provides another method of approximating the minimum mean square error V. This lower bound, which we denote by V_c^-, is given by

$$V_c^- = \left(I + \sigma^{-2}\right)^{-1} \tag{3-17}$$

where I is the Fisher information given by

$$I = E\left\{ p^{-2} \left(\frac{\partial p}{\partial x} \right)^2 \right\} \tag{3-18}$$

and p is the conditional density of the measurement y given the state x:

$$p = p(y|x) \quad .$$

In this simple example, p is the Gaussian sum distribution

$$p(y|x) = (1-\varepsilon)\psi_0(\xi) + \varepsilon\ \psi_1(\xi) \tag{3-19}$$

where

$$\begin{aligned} \xi &= y - x \quad , \\ \psi_i(\xi) &= (2\pi\ \sigma_1{}^2)^{-\ 1/2}\ \exp\{-\ 1/2\ \sigma_i{}^{-2}\ \xi^2\} \quad . \end{aligned} \tag{3-20}$$

It is easy to see that the Fisher information is given exactly by the integral

$$I = \int_{-\infty}^{\infty} \frac{\xi^2[(1-\varepsilon)\sigma_0{}^{-2}\ \psi_0(\xi) + \varepsilon\ \sigma_1{}^{-2}\ \psi_1(\xi)]^2}{(1-\varepsilon)\psi_0(\xi) + \varepsilon\ \phi_1(\xi)}\ d\xi \quad . \tag{3-21}$$

This is the same type of integral that occurred in the exact calculation of V. Although we cannot express this integral in closed-form, we can approximate it in the same way we approximated V. Thus, we find that

$$I = \sigma_0{}^{-2} - \varepsilon\ \sigma_0{}^{-4}(\sigma_1{}^2 - \sigma_0{}^2) + \varepsilon^2(\sigma_1{}^{-2} - \sigma_0{}^{-2})^2 K \tag{3-22}$$

where

$$K = \int_{-\infty}^{\infty} \frac{\xi^2\ \psi_1(\xi)^2}{(1-\varepsilon)\psi_0(\xi) + \varepsilon\ \psi_1(\xi)}\ d\varepsilon \quad .$$

Just as before (Eqs. 3-9 and 3-11) we can derive an upper and lower bound of K:

$$0 < K < \varepsilon^{-1}\ \sigma_1{}^2 \quad .$$

Thus, we obtain the corresponding bounds on V_c^-:

$$V_c^- \leq \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} - \frac{\varepsilon \dfrac{\sigma^4[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2}}{1 - \varepsilon \dfrac{\sigma^2[\sigma_1^2 - \sigma_0^2]}{\sigma_0^2[\sigma_0^2 + \sigma^2]}} \tag{3-23}$$

$$V_c^- \geq \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} - \frac{\varepsilon \dfrac{[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} \cdot \dfrac{\sigma_0^2 \sigma^4}{\sigma_1^2}}{1 - \varepsilon \dfrac{[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]} \dfrac{\sigma^2}{\sigma_1^2}} \tag{3-24}$$

From the inequality Eq. 3-23 we also see that

$$V_c^- = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} + o(\varepsilon) \tag{3-25}$$

where $o(\varepsilon)$ is an error term for which

$$\lim_{\varepsilon \to 0} \varepsilon^{-1} o(\varepsilon) = 0 \quad .$$

Thus, we see that the first order in ε, V_c^- agrees with V_1^+, the upper bound of Eq. 3-10 and also the first order expansion of V (as seen from Eq. 3-8). This suggests that V_c^- might be an accurate estimate of V, if we could compute V_c^- exactly. Note that theory tells us that $V_1^+ \geq V \geq V_c^-$. Thus, V_1^+ is, in fact, an upper bound of V_c^- and it is tighter than the upper bound in Eq. 3-23 as one can easily see.

One could develop an approximation of the integral K appearing in Eq. 3-22 and so develop second-order upper and lower bounds of I and thus V_c^- in the same way we do for V in Appendix A. We will not do so, but let us note that such analytic approximations of the Fisher information in Gauss-Markov hybrid state estimation problems are difficult to obtain even for simple problems. In Section 4 we will consider the Cramer-Rao lower bound for more general, dynamic hybrid state problems and find that there are good reasons to work to develop better methods to approximate the Fisher information.

3.4 RATE DISTORTION LOWER BOUND

The last method we will consider is based on rate distortion theory [16], a branch of information theory. Rate distortion theory allows us to derive an analytic, closed-form lower bound V_R^- of the minimum mean square error V. The basic approach is straightforward. The theory [16] tells us that the minimum mean square error V satisfies the inequality

$$R(V) \leq I(y;x)$$

where R(V) is the rate distortion function of the state x and I(y;x) is the mutual information between x and the measurement y. For a scalar Gaussian state, the rate distortion is simply

$$R(V) = \frac{1}{2} \ln\left(\frac{\sigma^2}{V}\right) .$$

This gives the lower bound

$$V \geq \sigma^2 \exp\{-2\, I(y;x)\} .$$

We cannot compute the mutual information I(y;x) exactly, but we can find a simple upper bound I^+ of I(y;x) and thus obtain a lower bound of V, namely

$$V > V_R^- = \sigma^2 \exp\{-2 I^+\} \quad . \tag{3-26}$$

To find an upper bound of mutual information, note that*

$$I(y;x) = h(y) - h(y|x) \tag{3-27}$$

where $h(y)$ is the differential entropy of the random variable y and $h(y|x)$ is the conditional differential entropy of y given x. The measurement y is a sum of the state x and the independent noise v which has a probability density

$$p(\eta) = \varepsilon\, p_1(\eta) + (1-\varepsilon) p_0(\eta)$$

where

$$p_i(\eta) = (2\pi\, \sigma_i^2)^{-1/2} \exp\{-1/2\, \eta^2\, \sigma_i^2\} \quad .$$

Because x and v are independent,

$$h(y|x) = h(v) \quad .$$

The differential entropy $h(v)$ is , by definition,

$$h(v) = \int_{-\infty}^{\infty} p(y)\ \ln p(y)\, dy \quad .$$

The function $f(p) = -p \ln p$ is concave in p, and application of Jenson's inequality yields the inequality

$$h(v) > \varepsilon\, h(v_1) + (1-\varepsilon) h(v_0)$$

where v_i has density p_i. Because v_i is Gaussian,

$$h(v_i) = \frac{1}{2} \ln(2\pi e\, \sigma_i^2) \quad .$$

*See [15] or [16] for the basic definitions and results of information theory.

Thus, we have the lower bound

$$h(y|x) \geq \frac{1}{2}\,\varepsilon\,\ell n(2\pi e\ \sigma_1^2) + \frac{1}{2}\,(1-\varepsilon)\ell n(2\pi e\ \sigma_0^2) \quad . \tag{3-28}$$

Another basic result of information theory gives us an upper bound on the differential entropy of a random variable in terms of its variance. In particular, we have

$$h(y) \leq \frac{1}{2}\,\ell n(2\pi e\ \Lambda) \tag{3-29}$$

where Λ is the variance of y, given by

$$\Lambda = \sigma^2 + \varepsilon\ \sigma_1^2 + (1-\varepsilon)\sigma_0^2 \quad .$$

Substituting this expression in Eq. 3-29 and using it together with Eq. 3-28 in Eq. 3-27, we obtain the upper bound I^+

$$I(y;x) \leq I^+ = \frac{1}{2}\,\ell n(2\pi e[\sigma^2 + \varepsilon\ \sigma_1^2 + (1-\varepsilon)\sigma_0^2]) - \frac{1}{2}\,\varepsilon\,\ell n(2\pi e\ \sigma_1^2) - \frac{1}{2}\,(1-\varepsilon)\ell n(2\pi e\ \sigma_0^2) \quad .$$

Using this expression for I^+ in Eq. 3-26 gives us the rate distortion lower bound

$$V_R^- = \frac{\sigma^2\ \ \sigma_1^2\varepsilon\ \ \sigma_0^2(1-\varepsilon)}{\sigma^2 + \varepsilon\ \sigma_1^2 + (1-\varepsilon)\sigma_0^2} \tag{3-30}$$

Note that to first order in ε, the bound V_R^- is given by

$$V_R^- = \frac{\sigma^2\ \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon\ \frac{\sigma^2\ \sigma_0^2}{[\sigma_0^2 + \sigma^2]^2}\left\{2\ \ell n\left(\frac{\sigma_1}{\sigma_0}\right)[\sigma_0^2 + \sigma^2] - [\sigma_1^2 - \sigma_0^2]\right\} + O(\varepsilon) \quad . \tag{3-31}$$

3.5 NUMERICAL EXAMPLE

Figure 3-1 shows the mean square error predictions made by the different performance analysis methods discussed in this section. The statistical parameters of the variables in Eq. 3-1 were chosen so that

$$\bar{x} = 0 ,$$

$$\sigma^2 = 100 ,$$

$$\sigma_1^2 = 100 ,$$

$$\sigma_0^2 = 1 .$$

The false alarm probability $\varepsilon = P\{q=1\}$ was varied from 0 to .5.

Figure 3-1 shows performance predictions obtained from a direct first-order expansion (an upper bound V_1^+ in Eq. 3-10 and a lower bound V_1^- in Eq. 3-12), a direct second-order expansion (an upper bound V_2^+ and a lower bound V_2^- in Eq. A-25 of Appendix A), and a rate distortion bound (a lower bound V_R^- in Eq. 3-30). Note that the direct first-order expansion V_1^+ is equal to the first-order expansion of the Cramer-Rao lower bound V_C^-. Mean square error in Fig. 3-1 is normalized by dividing by the initial variance σ^2.

For small ε the direct second-order expansion should give tight upper and lower bounds of mean square error. As can be seen in Fig. 3-1, for $\varepsilon > .3$ the second-order expansion becomes less accurate than the first-order expansion.

Note that the rate distortion lower bounds for this example is less accurate than the other performance predictions. This appeared to be true in all numerical examples we considered.

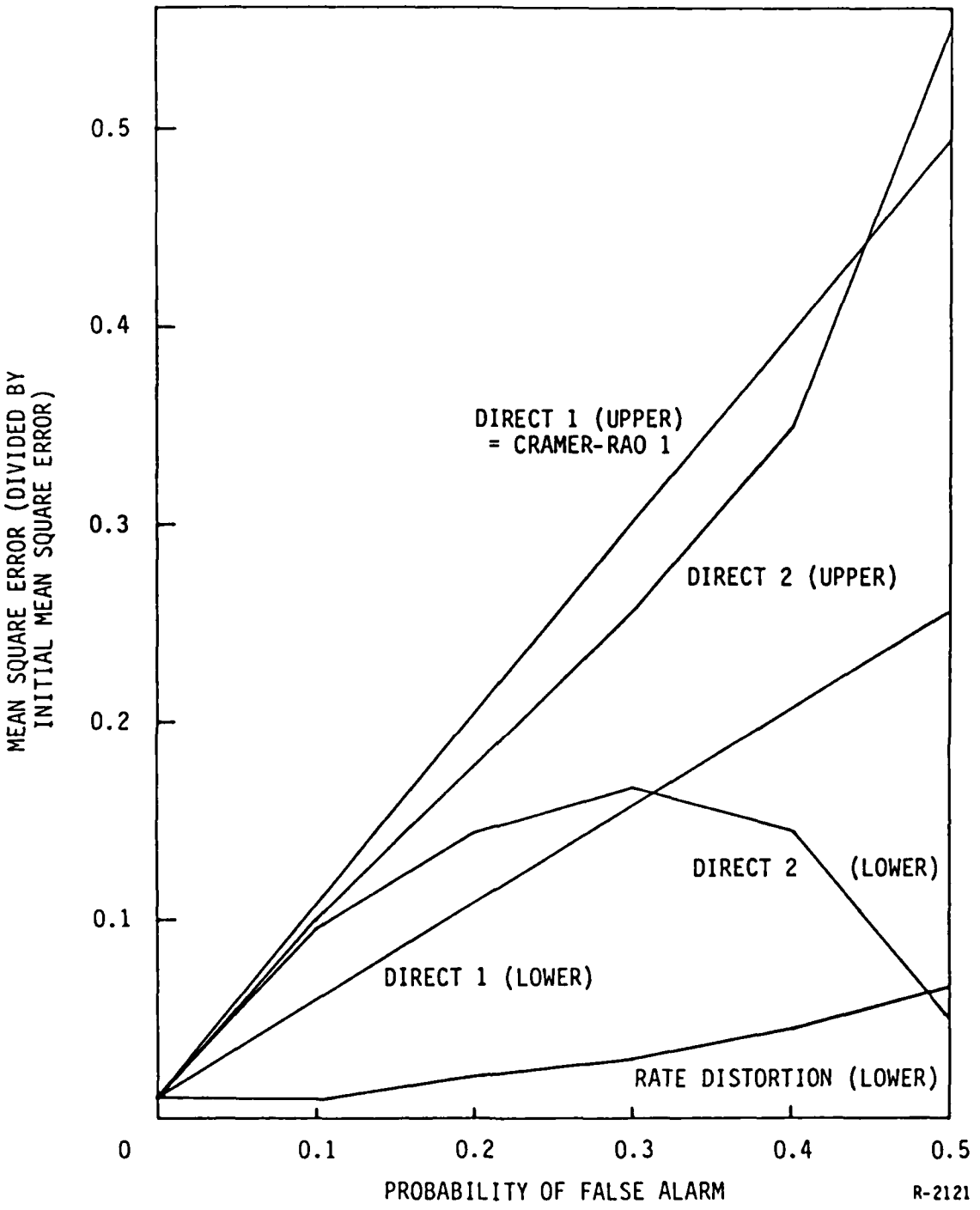


Figure 3-1. Mean Square Performance Bounds

3.6 REMARKS

In this section we have considered several performance analysis techniques applied to a simple static Type I hybrid state estimation problem. Exact closed-form evaluation of minimum mean square error is impossible to achieve even for this simple problem. Typically, one encounters integrals which have as integrands the ratio of two Gaussian sum distributions. These integrals appear in the exact error, the Cramer-Rao lower bound, and in the second-order asymptotic approximation (Appendix A). Approximating these integrals is difficult even for the simple problem of this section. Generalization to dynamic problems of interest seems unlikely.

On the other hand, the rate distortion lower bound is easy to compute, but its accuracy seems very poor compared to the other approaches described in this section. The first-order asymptotic approximations are both easy to compute and reasonably accurate. Unfortunately, this type expansion does not appear to generalize to dynamic problems of interest.

In the following sections we will consider performance analysis techniques for more general dynamic hybrid state estimation problems. Section 4 discusses a Cramer-Rao approach for Type I problems, Section 5 discusses a rate distortion approach for Type II problems, and Section 6 discusses a Monte-Carlo approach that potentially applies to any hybrid state problem.

SECTION 4

CRAMER-RAO LOWER BOUND FOR TYPE I PROBLEMS

4.1 INTRODUCTION

Because Type I problems can be formulated as a partially observed Markov process $\langle x(t), y(t) \rangle$ for which $x(t)$ is continuous, it is theoretically possible to apply classical performance analysis techniques to such problems. In this section we will examine the applicability of Cramer-Rao methods to Type I problems, and particularly to the subclass of Type I problems described by Eq. 2-9.

The Cramer-Rao method as presented in Van Trees [4] provides a lower bound of the estimation error covariance (for which lower is interpreted in the sense of symmetric matrix inequalities) in Bayesian estimation problems in which both state and measurement are modeled as random variables. Thus, the Cramer-Rao method gives us an estimate of mean square error. Although the bound is given for static problems in [4], it has been extended to various dynamic filtering problems in [5]-[7] (Bayesian framework) and [8] (non-Bayesian framework). The method is useful because it gives us a computable performance measure for a large class of nonlinear estimation problems. In particular, the computational complexity of the bound for the filtering problems described in [5]-[8] is comparable to that required to compute the error covariance for a linear Gaussian filter and is proportional to the number of time periods under investigation. This is significantly less complexity than that of the optimal filters for most nonlinear problems.

We will show here that the Cramer-Rao method applies similarly to Type I hybrid state estimation problems described by Eq.2-9. Actually, we show more generally that this method applies to any estimation problem for which the state process satisfies Eq. 2-9a and the measurement process y(t) is such that all y(t), t=1,2,..., are conditionally independent and identically distributed given the state process. We also assume that the conditional distribution of y(t) depends just on x(t) and that this distribution p(y(t)|x(t)) is a continuously differentiable function of x(t) which vanishes as |x(t)| tends to infinity.

Let J(t) denote the information matrix given by

$$J(t) = E\{p(y(t)|x(t))^{-2} p_x(y(t)|x(t))^T p_x(y(t)|x(t))\} \qquad (4\text{-}1)$$

where E{ } denotes taking the expectation with respect to the joint measurement and state processes, $p_x(y|x)$ is the partial derivative of $p(y|x)$ with respect to x (a row vector if x is a column vector), and p_x^T denotes the transpose of p_x (a column vector if p_x is a row vector). In terms of J(t) the Cramer-Rao lower bound on the average error covariance P(t) is given by the linear filtering update equations, namely

$$P(t+1) = [(AP(t)A^T + Q)^{-1} + J(t)]^{-1} \qquad (4\text{-}2)$$

where Q is the covariance of the process noise w(t) in Eq. 2-9a. We will present only a sketch of the proof here. One expresses the filtering problem of Eq. 2-9 as a large dimensional static estimation problem of estimating the entire state trajectory X given the measurement trajectory Y. Using the static result of [4], note that the Cramer-Rao bound for this problem is equal to the bound (indeed, equal to the exact error covariance) for a linear

Gaussian estimation problem for which Eq.2-9a holds and Eq. 2-9b is replaced by the linear equation

$$y(t) = C(t)x(t) + v(t) \tag{4-3}$$

in which $C(t)$ and the covariance $R(t)$ of $v(t)$ are chosen so that

$$J(t) = C(t)^T R(t)^{-1} C(t) \quad . \tag{4-4}$$

This is always possible for any information matrix $J(t)$ (note that it is not necessary to compute $C(t)$ or $R(t)$). It follows that the filter estimation error covariance of the hybrid state problem is bounded below by the corresponding error covariance for the linear problem given by Eqs. 2-9a and 4-3, and therefore, the bound can be computed according to Eq. 4-2.

4.2 COMPUTATION OF INFORMATION MATRIX FOR GAUSSIAN MIXTURES

From above we see that applying the Cramer-Rao approach depends on computing the information matrix $J(t)$ in Eq. 4-1 for static problems of the form of Eq. 2-9b, namely

$$y = C(q)x + D(q)v \quad , \tag{4-5}$$

where x and v are Gaussian distributed and q takes the discrete values $1,2,\ldots,N$ with probabilities b_k for each value $q = k$. We will assume without loss of generality that v has 0 mean and an identity covariance matrix. Let R_k denote the covariance matrix of $D(k)v$, that is $R_k = D(k)D(k)^T$. In this case, the conditional probability distribution $p(y|x)$ needed in Eq. 4-1 is

$$p(y|x) = \sum_{k=1}^{N} b_k p_k(y|x) \tag{4-6}$$

where p_k is the Gaussian density

$$p_k(y|x) = (\det 2\pi R_k)^{-1/2} \exp\left(-1/2[y-C(k)]^T R_k^{-1}[y-C(k)]\right) . \quad (4-7)$$

The corresponding derivative $p_x(y|x)$ is

$$p_x(y|x) = \sum_{k=1}^{N} b_k[y-C(k)]^T R_k^{-1} C(k) p_k(y|x) . \quad (4-8)$$

Define J(x) for each x as

$$J(x) = E\{p(y|x)^{-2} p_x(y|x)^T p_x(y|x)|x\} \quad (4-9)$$

where the expectation $E\{\ |x\}$ is conditioned on x. Thus, the information matrix J is found by integrating J(x) with respect to the Gaussian distributed x.

The expectation in Eq. 4-9 involves integrating the ratio of two sums of Gaussian densities given by substituting Eq. 4-7 and 4-8. In general it is not possible to do this integration in closed form (in terms of elementary functions), nor is J(x) integrable with respect to the Gaussian density of x in closed form. Thus, one must resort to some type of numerical integration to evaluate J(x) and J. Since the integrals are multidimensional and have infinite domains of integration, we have chosen to use the control variate Monte-Carlo method of approximating integrals [9],[10].

4.3 REMARKS ON CRAMER-RAO APPROACH

The advantage of the Cramer-Rao approach is that the complexity of the computation is essentially limited to the complexity of a static problem depending on the dimensions of the state vector x(t), the measurement vector y(t), and the number of discrete states q(t). Once this static problem is

solved (namely the computation of J(t) in Eq. 4-1), the dynamic problem is easily solved by the familiar recursion in Eq. 4-2.

The disadvantage of the Cramer-Rao approach for hybrid state performance analysis is that the information matrix J(t) of Eq. 4-1 may be difficult to compute even for small dimensional static problems. In general this computation requires integrating a ratio of two sums of Gaussian density functions, and thus, the integral has no closed form expression in terms of elementary functions. This is in marked contrast with the situation for nonlinear filtering problems in Gaussian white noise where J(t) can be computed in closed form for a large class of nonlinearities [11],[12]. As noted above, we have employed Monte-Carlo methods to compute these integrals, but even these methods become very computationally expensive if the static dimensions of the problem are large. This would be the case for realistic multiobject tracking problems with many targets and many returns per measurement period.

The Cramer-Rao method produces a lower bound to the minimum mean square estimation error; however, this lower bound may be much lower than the minimum mean square error -- that is, the bound can be very loose. In principle, the bound becomes tight as the measurement covariance tends to zero. More precisely, the conditional covariance of y in Eq. 4-5 given x must be small. In practice, one must determine how small by numerical computation and comparison to more accurate performance measures.

We wish to mention two previous applications of Cramer-Rao methods to hybrid state estimation problems. If the measurement is purely discrete or is the sum of a continuous and discrete component, one can compute Cramer-Rao type bounds fairly easily in a large number of cases [5],[13]. In fact, in this case it is possible to deal with a discrete process q(t) that is dependent

on the continuous process x(t). These computations depend crucially on the fact that the discrete state enters only additively in the measurement equation. The multiplicative dependence we have assumed severely complicates the problem.

The work reported in [14] is much closer to the problems presented here. That work studies a specific hybrid state parameter estimation problem using the Cramer-Rao method for unknown, nonrandom parameters (see [4]). The approach in [14] is somewhat different as the state x(t) is not random as it is here, and the approach is a batch procedure rather than the recursive procedure presented here. Furthermore, uniformly distributed continuous input noise variables are used while we only use Gaussian distributed input noise.

SECTION 5

RATE DISTORTION LOWER BOUNDS

5.1 RATE DISTORTION METHOD

Information theory provides another approach to bounding estimation performance based on Shannon's theory of rate distortion [15],[16]. The basic mathematical idea behind that approach is that the mean square error D of any estimator of a random variable X given measurement Y is constrained by the inequality

$$R_X(D) \leq I(X;Y) \tag{5-1}$$

where R_X is the rate distortion function of X and I(X;Y) is the mutual information between X and Y [16]. Solving Eq. 5-1 for the largest D satisfying the inequality gives a lower bound of the minimum mean square error of estimating X given the measurement Y. In a filtering problem one would let X be the current state component $x_i(t)$ of interest and let Y be the measurement sequence $\langle y(t),y(t-1),\ldots,y(1)\rangle$ up to the current time. This method has been applied to nonlinear filtering problems with continuous states and measurements generated by Gaussian noise inputs in [17]-[20]. We will apply the method to Type I and II hybrid state estimation problems in this section.

5.2 RATE DISTORTION BOUND FOR TYPE II PROBLEMS

The difficulty of applying the rate distortion method depends on the difficulty of calculating the rate distortion function R_X and the mutual

information I(X;Y) in Eq. 5-1. In general (and this is true for the hybrid state estimation problem), these expressions cannot be computed in closed form. Therefore, we look for an upper bound I_U of the mutual information and a lower bound R_L of the rate distortion function. The Eq. 5-1 is replaced by

$$R_L(D) < I_U \quad . \tag{5-2}$$

Solving Eq. 5-2 for the largest D that satisfies the inequality gives a lower bound of the minimum mean square estimation error which is lower than the bound given by Eq. 5-1.

Let us first find a lower bound for the rate distortion function. Shannon's lower bound of the rate distortion function R_X is

$$R_X(D) > h(p_X) - 1/2 \log(2\pi\, eD) \tag{5-3}$$

where $h(p_X)$ denotes the differential entropy of the probability density p_X of X [16]. To avoid notational complexity but without loss of generality (as will become evident), let x(t) be a scalar process. The density $p_x(t)$ is a Gaussian mixture, and although we cannot compute $h(p_x(t))$, we can find a lower bound for it, namely

$$h(p_x(t)) > \sum_q h(p_{x(t)|q})\, p(q,t) \tag{5-4}$$

where q denotes the discrete sequence $\langle q(t),\ldots,q(1)\rangle$, p(q,t) denotes the probability of this sequence occurring, and $p_{x(t)|q}$ is the Gaussian density of x(t) conditioned on the sequence q. This gives us a lower bound of the rate distortion function. For reference below let P(t|q) denote the conditional covariance of x(t) given the sequence q.

Next we find an upper bound for the mutual information. The basic relation is

$$I(x(t);y(t),\ldots,y(1)) \leq I(x(t),\ldots,x(1),q(t),\ldots,q(1);y(t),\ldots,y(1)) \tag{5-5}$$

The right side of Eq. 5-5 is given by

$$\begin{aligned} &I(x(t),\ldots,x(1),q(t),\ldots,q(1);y(t),\ldots,y(1)) \\ &= h(y(t),\ldots,y(1)) \\ &- h(y(t),\ldots,y(1)|x(t),\ldots,x(1),q(t),\ldots,q(1)) \end{aligned} \tag{5-6}$$

where the expression $h(Y|X)$ denotes the conditional differential entropy of Y with respect to X [16]. Let $M(t|q)$ denote the covariance matrix of the measurement sequence $\langle y(t),\ldots,y(1)\rangle$ conditioned on knowing the discrete sequence $q = \langle q(t),\ldots,q(1)\rangle$. Then the differential entropy is bounded above by

$$h(y(t),\ldots,y(1)) \leq 1/2 \log\left[\det\left(2\pi e \sum_{q} p(q,t)M(t|q)\right)\right] \tag{5-7}$$

and the conditional differential entropy is given exactly by

$$\begin{aligned} &h(y(t),\ldots,y(1)|x(t),\ldots,x(1),q(t),\ldots,q(1)) \\ &= 1/2 \sum_{q} p(q,t) \log\left[\det\left(2\pi e R(t|q)\right)\right] \end{aligned} \tag{5-8}$$

where $R(t|q)$ denotes the covariance matrix of the noise sequence $D(q(t))v(t)$ given the discrete sequence $q(t)$. Eqs. 5-7 and 5-8 give us an upper bound of the mutual information.

Using the lower bound of the rate distortion function and the upper bound of the mutual information, we obtain the following lower bound of the mean square estimation error D_t of estimating $x(t)$ given the measurements up to time t.

$$D_t > \exp(F_t + G_t - H_t) \tag{5-9}$$

where F_t, G_t, and H_t are given by

$$F_t = \sum_q p(q,t)\log\left(P(t|q)\right) \tag{5-10}$$

$$G_t = \sum_q p(q,t)\log\left(\det[R(t|q)]\right) \tag{5-11}$$

$$H_t = \log\left(\det[\sum_q p(t,q)M(t|q)]\right) \tag{5-12}$$

Note that if the noise random variables v(t) are independent, standard Gaussian random vectors, then R(t|q) is the block diagonal matrix

$$R(t|q) = \text{diag}\left\{D(q(k))\ D(q(k))^T\right\} \tag{5-13}$$

and

$$\det R(t|q) = \prod_{k=1}^{t} \det\left\{D(q(k))\ D(q(k))^T\right\} \quad . \tag{5-14}$$

If the q(k) are independent and identically distributed, then

$$G_t = t \cdot E\left\{\log \det\left(D(q(1))\ D(q(1))^T\right\} \quad . \tag{5-15}$$

5.3 REMARKS ON RATE DISTORTION METHOD

Note that no assumption was made about the statistics of the discrete state process q(t) except that they are independent of the continuous states. Thus, the rate distortion bound allows us to treat Type II problems in which q(t) is a finite state Markov chain. In theory, one could treat more general problems (Type III) using the rate distortion theory approach provided that

one can compute or bound the corresponding rate distortion function for the state component and the performance criterion of interest. For example, one can do this if one is interested in just estimating discrete states and the average probability of error criterion is used. See [16] for the general theory of rate distortion with general error criteria.

The rate distortion method has a different kind of computational complexity from the Cramer-Rao method presented above. One can compute this bound exactly (i.e., no numerical integrations are required), but the complexity increases exponentially as the number of time periods t considered. The computations are similar in nature to the computations occurring in the optimal hybrid state estimation algorithm [1],[2] -- that is, for each sequence of discrete states one computes $M(t|q)$ and $P(t|q)$ from the corresponding linear filter equations. Of course, these computations are exact and no simulation or Monte-Carlo averaging is involved. Still, if there are N discrete values of $q(t)$, then this computation will involve N^t filter computations to consider a time period of t steps. Furthermore, the covariance $M(t|q)$ has dimension mt where m is the dimension of $y(t)$ and t is the number of time periods considered. Computing the determinants needed in Eqs. 5-11 and 5-12 requires on the order of $(mt)^3$ operations (see [20],[21] for an efficient method for doing this in the context of rate distortion problems). Thus, although the rate distortion method is easier than the Cramer-Rao method to compute for short periods of time, the rate distortion computation quickly overtakes the Cramer-Rao in complexity as the number of time periods increases. This method has a similar sensitivity to the number of discrete states and the dimensions of the state and measurement vectors.

The rate distortion method gives a lower bound of minimum mean square error, and in some cases that lower bound may be very loose. In principle, the bound becomes tight as the measurement covariance increases, but in practice (as with the Cramer-Rao method) one has to check the bound with more accurate performance measures computed for some simple test cases in order to get a feel for the rate distortion bound's accuracy.

SECTION 6

MONTE-CARLO PERFORMANCE ANALYSIS BASED ON THE REPRESENTATION THEOREM

6.1 INTRODUCTION

A major disadvantage of the rate distortion method described in Section 5 is that it can be inaccurate unless measurement noise is very high. In [28] we faced a similar problem in applying rate distortion methods to nonlinear filtering problems such as occur in passive tracking problems. In this section we will extend a Monte-Carlo method discovered in [28] to include hybrid state estimation problems. Note that this method is based on a rate distortion approach due to Galdos [18],[19]. This method uses the representation theorem of nonlinear filtering theory to obtain an approximation of a rate distortion lower bound on estimation error. Galdos presented both a perturbation approximation and a Monte-Carlo approximation for his method in [18]. In [19] he clarified, extended, and simplified the Monte-Carlo approach. In this section we show that one can use the Monte-Carlo approach directly to approximate the minimum mean square estimation error without invoking any rate distortion theory. As we will see, this direct method not only avoids the lower bound aspect of rate distortion theory, but is also is simpler to implement and more generally applicable than the indirect rate distortion approach of Galdos. In subsection 6.2 we will generalize Galdos' method to general hybrid state systems. In subsection 6.3 we will present our direct method, and in subsection 6.4 we will make some concluding remarks.

6.2 GALDOS RATE DISTORTION METHOD

The problem of [19] is that of estimating optimal performance in discrete time nonlinear filtering problems described by equations of the form

$$x(t+1) = f(t,x(t)) + g(t,x(t))w(t) \tag{6-1}$$

$$y(t) = h(t,x(t)) + v(t) \tag{6-2}$$

where $w(t)$ and $v(t)$ are zero mean Gaussian white noise sequences. In particular, if $d(x(t))$ is a scalar function of the state $x(t)$, it is desired to approximate the minimum mean square error for estimating $d(x(t))$ given measurements $y(s)$ for $s = 1,\ldots,t$.

The method of [19] is not limited to systems of the form of Eqs. 6-1 and 6-2. As we will show, one can consider any problem such that the measurements $y(s)$ are conditionally independent given the state process $x(s)$, $1 \leq s \leq t$, and the conditional density that $y(s) = y$ given $x(s) = x$ is $p(y|x,s)$. One also needs to be able to simulate independent samples of measurement and state processes — we will explain more precisely what is needed below.

Rate Distortion Approach [16],[19]

One can consider the problem of estimating the minimum expected value of $\rho(d(x(t)),\hat{d}(t))$ where ρ is some arbitrary error criterion. The minimum is taken over all estimates $\hat{d}(t)$ which depend only on the measurements $y(1),\ldots,y(t)$. The rate distortion approach is based on the inequality

$$R(\varepsilon_t) \leq I(d(x(t));Y_t) \tag{6-3}$$

where

$R(D)$ = rate distortion function with respect to ρ;

$\varepsilon_t = E\{\rho(d(x(t)),\hat{d}(t))\}$;

$\hat{d}(t)$ = estimate of $d(t)$ based on Y_t;

Y_t = measurement sequence $y(1),\ldots,y(t)$;

ρ = error criterion;

I = mutual information between $d(x(t))$ and Y_t.

To use Eq. 6-3 you must be able to compute R (or rather its inverse function R^{-1}) and the mutual information $I(d(x(t));Y_t)$. Then

$$\varepsilon_t \geq R^{-1}(I(d(x(t));Y_t)) \quad .$$

If we cannot compute the rate distortion function or the mutual information, then we must approximate by means of a lower bound on the rate distortion function,

$$R_L(D) \leq R(D)$$

and an upper bound on the mutual information.

$$I_U(t) \geq I(d(x(t));Y_t) \quad .$$

This gives the looser inequality

$$\varepsilon_t \geq R_L^{-1}(I_U(t)) \tag{6-4}$$

for the estimation error.

Galdos uses the usual lower bound R_L on the mean square distortion function R for the error criterion $\rho(d,\hat{d}) = (d-\hat{d})^2$ and the source $d(x(t))$ given by

$$R_L(D) = h(p_{d(x(t))}) - \frac{1}{2}\log(2\pi eD) \tag{6-5}$$

where $p_{d(x(t))}$ denotes the probability density of $d(x(t))$ and $h(p)$ denotes the differential entropy of a probability density p. Note that R_L (and R also) depends on t implicitly through this differential entropy of $d(x(t))$. Note that if d is a linear functional of $x(t)$ and if $x(t)$ is Gaussian distributed (as it is for the subclass of Type I hybrid state systems described in Eq. 2-9, for example), then the differential entropy is easily computed as a function of the variance V of $d(x(t))$, namely

$$h(p_{d(x(t))}) = \frac{1}{2} \log(2\pi eV) \quad . \tag{6-6}$$

The mutual information is not so easy to approximate, and this is where Galdos resorts to Monte-Carlo approximation. By definition, the mutual information is

$$I(d(x(t));Y_t) = E\left\{\log\left\{\frac{p_{y|d}\{Y_t|d(x(t))\}}{p_y\{Y_t\}}\right\}\right\} \quad . \tag{6-7}$$

In the right-hand side expression of Eq. 3-7, we use $p_{y|d}$ to denote the conditional density of the measurement sequence Y_t given $d(x(t))$, and we use p_y to denote the unconditional density of Y_t.

Representation Theorem

The representation theorem (for discrete-time estimation, this is just Bayes' rule) is used to represent the ratio

$$\frac{p_{y|d}\{Y_t|d(x(t))\}}{p_y\{Y_t\}} = \frac{E_{x|d}\{\exp(\zeta_t)|d(x(t))\}}{E_x\{\exp(\zeta_t)\}} \tag{6-8}$$

where ζ_t is the random variable defined by

$$\zeta_t(Y_t,X_t) = \sum_{s=1}^{t} \log\left[p\left(y(s)|x(s),s\right)\right] \quad . \tag{6-9}$$

Thus, ζ_t is a function of the measurement sequence Y_t and the state sequence $X_t = (x_1,\ldots,x_t)$. The expectations $E_{x|d}$ and E_x integrate over the state sequence only, leaving the measurement sequence alone. To be precise, let p_x denote the probability density of the state sequence X_t. Then E_x is interpreted to mean

$$E_x\{\exp(\zeta_t)\} = \int \exp\left(\zeta_t(Y_t,X_t)\right)\, p_x\{X_t\}\, dX_t \quad . \tag{6-10}$$

Thus, this expectation produces a function of Y_t. If $p_{x|d}$ denotes the conditional density of the state sequence given the variable $d(x(t))$, then $E_{x|d}$ is interpreted to mean

$$E_{x|d}\{\exp(\zeta_t)\} = \int \exp\left(\zeta_t(Y_t,X_t)\right)\, p_{x|d}\{X_t|d(x(t))\}\, dX_t \quad . \tag{6-11}$$

This expectation produces a function of Y_t and $d(x(t))$.

Indirect Monte-Carlo Approximation

The approach of [19] is to approximate the integrals in Eqs. 6-10 and 6-11 by Monte-Carlo methods, and then use Monte-Carlo methods again on the approximate ratio corresponding to Eq. 6-8 to obtain the final approximation for the mutual information in Eq. 6-7. We will now outline a simple Monte-Carlo algorithm similar to [19] for approximating the mutual information. Note that the integers n and m control the number of simiulations used in the Monte-Carlo approximation.

Step 1. Initialize $I_t \leftarrow 0$ and $i \leftarrow 1$.

Step 2. Simulate $d(x(t))$ and the measurement sequence $y(1),\ldots,y(t)$ from the joint state and measurement model.

Step 3. Initialize $N_t \leftarrow 0$, $D_t \leftarrow 0$, and $j \leftarrow 1$.

Step 4. Simulate the state sequence $x'(s)$, $1 \le s \le t$, conditioned on $d(x'(t)) = d(x(t))$.

Step 5. $$N_t \leftarrow \frac{j}{j+1} N_t + \frac{1}{j+1} \exp\left[\sum_{s=1}^{t} \log\left[p(y(s)|x'(s),s)\right]\right] .$$

Step 6. Simulate a new state sequence $x'(s)$, $1 \le s \le t$, without conditioning.

Step 7. $$D_t \leftarrow \frac{j}{j+1} D_t + \frac{1}{j+1} \exp\left[\sum_{s=1}^{t} \log\left[p(y(s)|x'(s),s)\right]\right] .$$

Step 8. If $j = m$, then go to next step, else $j \leftarrow j+1$ and return to Step 4.

Step 9. $$I_t \leftarrow \frac{i}{i+1} I_t + \frac{1}{i+1} \log\left[\frac{N_t}{D_t}\right] .$$

Step 10. If $i = n$, then stop, else $i \leftarrow i+1$ and return to Step 2.

When the algorithm stops, the value of I_t will be an approximation of the mutual information in Eq. 6-7. Using this with Eq. 6-5 in the inequality of Eq. 6-4 gives the approximate lower bound

$$\varepsilon_t \ge \frac{1}{2\pi e} \exp\{2[h(p_{d(x(t))}) - I_t]\} \qquad (6\text{-}12)$$

for the mean square error ε_t of estimating $d(x(t))$.

Remarks on Indirect Monte-Carlo Approach

A few remarks are in order at this point. The algorithm consists of an outer loop that iterates n times (Steps 2 through 10) and an inner loop that iterates m times (Steps 4 through 8) for each iteration of the outer loop. Thus the algorithm requires of the order of nm simulations (it requires n

unconditional simulations of the state and measurement sequences, nm unconditional simulations of the state sequence, and nm conditional simulations of the state sequence). As $n,m \to \infty$, the approximation I_t converges to the mutual information. However, the rate of convergence is not clear. For some numerical experiments, n and m must approach 1000 for reasonable accuracy. Thus, the simple Monte-Carlo approach outlined above can be very time-consuming, although the algorithm is very simple to implement.

A more difficult problem is simulating the conditional state sequence in Step 4. Theoretically, this involves simulating a time reversed state x(t) process [19],[29] from a terminal condition specified on $d(x(t))$. If the state process is Gaussian and d is a linear functional of the state vector, then the reversed process can be simulated by a linear Gaussian model as outlined in [19] and derived in [29. If the state is non-Gaussian or nonlinear, or d is nonlinear, then the backward simulation problem is considerably more difficult. Galdos suggests an approximate approach in [19].

Computing the differential entropy $h(p_{d(x(t))})$ is easily accomplished by using Eq. 6-6 if the state process is Gaussian and d is linear. Note also that in the linear case the approximation R_L in Eq. 6-5 is exact: $R_L = R$. If one or the other is nonlinear, then the differential entropy may also have to be approximated.

Finally, the method is fundamentally limited by the fact that, at best, it is a lower bound on performance. There is no guarantee that the rate distortion lower bound will be a good approximation of the estimation error.

Despite these limitations, the method has some appeal because it potentially applies to a very general class of filtering problems and requires little special analysis once the basic algorithm is set up. In the next

section we will show how to use the representation theorem in a direct Monte-Carlo approach that avoids many of the problems inherent in the indirect rate distortion approach.

6.3 DIRECT MONTE-CARLO PERFORMANCE ANALYSIS

The direct method we propose is based on the observation that the representation theorem used in [19] essentially gives the conditional probability density π_t of $x(t)$, and this density allows one to compute the exact minimum mean square error ε_t in estimating $d(x(t))$ via the formula

$$\varepsilon_t = E\left\{ \int d(\xi)^2 \pi_t(\xi) d\xi - \left(\int d(\xi) \pi_t(\xi) d\xi \right)^2 \right\} \quad . \tag{6-13}$$

To see how this is done, consider the representation theorem of Eq. 6-8 again. Let $p_{x(t)}$ denote the unconditional density of $x(t)$ and let $E_{x|x(t)}$ denote the conditional integral defined in Eq. 6-11. Then the conditional density π_t is given by

$$\pi_t(\xi) = \frac{E_{x|x(t)}\{\exp(\zeta_t)|\xi\}}{E_x\{\exp(\zeta_t)\}} p_{x(t)}\{\xi\} \quad . \tag{6-14}$$

Thus, the conditional integrals in Eq. 6-13 can be expressed as

$$\int d(\xi)\pi_t(\xi)d\xi = \int d(\xi) \frac{E_{x|x(t)}\{\exp(\zeta_t)|\xi\}}{E_x\{\exp(\zeta_t)\}} p_{x(t)}\{\xi\} d\xi \quad , \tag{6-15}$$

which is same as

$$\int d(\xi)\pi_t(\xi)d\xi = \frac{E_x\{d(x(t))\exp(\zeta_t)\}}{E_x\{\exp(\zeta_t)\}} \quad . \tag{6-16}$$

Similarly, we have

$$\int d(\xi)^2 \pi_t(\xi) d\xi = \frac{E_x\{d(x(t))^2 \exp(\zeta_t)\}}{E_x\{\exp(\zeta_t)\}} \quad . \tag{6-17}$$

We will now use these formulas to derive a direct Monte-Carlo approach to approximating minimum mean square error.

Direct Monte-Carlo Algorithm

We describe a simple Monte-Carlo algorithm analogous to the one given for the rate distortion approach. As above, the integers n and m determine the number of simulations used by the Monte-Carlo approximation.

Step 1. Initialize $\varepsilon_t \leftarrow 0$, $i \leftarrow 1$.

Step 2. Simulate a measurement sequence $y(s)$, $1 \le s \le t$.

Step 3. Initialize $V_t \leftarrow 0$, $M_t \leftarrow 0$, $P_t \leftarrow 0$, $j \leftarrow 1$.

Step 4. Simulate the state sequence $x(s)$, $1 \le s \le t$.

Step 5. $\zeta_t \leftarrow \sum_{s=1}^{t} \log[p(y(s)|x(s),s)]$.

Step 6. $V_t \leftarrow \frac{j}{j+1} V_t + \frac{1}{j+1} d(x(t))^2 \exp(\zeta_t)$.

Step 7. $M_t \leftarrow \frac{j}{j+1} M_t + \frac{1}{j+1} d(x(t)) \exp(\zeta_t)$.

Step 8. $P_t \leftarrow \frac{j}{j+1} P_t + \frac{1}{j+1} \exp(\zeta_t)$.

Step 9. If $j = m$, then go to next step, else $j \leftarrow j+1$ and return to Step 4.

Step 10. $\varepsilon_t = \frac{i}{i+1} \varepsilon_t + \frac{1}{i+1} \left[\frac{V_t}{P_t} - \left[\frac{M_t}{P_t} \right]^2 \right]$.

Step 11. If $i = n$, then stop, else $i \leftarrow i+1$ and return to Step 2.

The final value of ε_t will be an approximation to the minimum mean square filtering error for estimating $d(x(t))$.

Remarks on Direct Monte-Carlo Approach

The direct Monte-Carlo approach outlined above also requires of the order of nm simulations (n simulations of the state and measurement sequences, and nm simulations of the state sequence alone). Furthermore, the direct method can converge slowly just as the indirect method, and n and m may have to be very large to obtain sufficient accuracy.

However, in other respects, the direct method is an improvement over the indirect method. The direct does not require conditional simulation of the state sequence; only forward time simulations are needed, nor does the method require any entropy calculations. Consequently, the direct method is easy to apply to problems with non-Gaussian noise and nonlinear state dynamics. In fact, the type of state model is irrelevant to the application of the algorithm, provided one can simulate the state sequence and provided that the measurement model is given the probability density function $p(y|x,s)$. Thus, it is possible to apply the direct method to very general non-Gaussian, non-Markovian state processes.

Finally, the direct method approximates the actual minimum mean square error and not a lower bound of this error. Thus, as the Monte-Carlo parameters n and m increase, one obtains better approximations of the optimal performance. For the indirect method, the performance approximation may be poor no matter how large n and m are.

6.4 CONCLUSIONS

In [18],[19] Galdos had the innovative idea of using the representation theorem of nonlinear filtering theory for computational rather than theoretical purposes to obtain a rate distortion lower bound of optimal filtering error approximations for a large class of filtering problems. In this section

we have shown how to apply the representation theorem directly without having to use rate distortion theory, thus obtaining a simpler, more accurate approximation of optimal filtering error which applies more easily to a larger class of hybrid state filtering problems.

The technique outlined in the previous subsection has the advantage that it is very easy to use for almost any filtering problem. Its major disadvantage is due to the slow convergence of the crude Monte-Carlo methods employed. This suggests a problem for further work. The paper [30] of LeGrand is of interest in studying this problem as he has investigated the problem of accelerating Monte-Carlo convergence for conditional expectation type calculations using importance sampling. Note also that the algorithm described in subsection 6.3 can be partly decomposed into parallel computations, so the algorithm is a good candidate for array processing.

Any simulation technique, such as the one described in this section, has the fundamental limitation that it only indicates performance for a single set of parameters defining the statistics of the underlying model. Thus, such methods cannot easily provide insight into how performance is affected by parametric variations. Unfortunately, accurate analytic methods which might provide such insight are lacking for most non-Gaussian, nonlinear filtering problems. The method discussed here offers a viable approach to accurate optimal performance analysis which may at least help determine when practical nonlinear filters are near-optimal. If the computational problems described above can be solved by algorithmic or computer processing improvements, it is also possible that the technique might give useful insights into hybrid state estimation comparable to what is possible with analytic techniques available for simpler problems.

SECTION 7

A RANDOM POINT PROCESS APPROACH TO MULTIOBJECT TRACKING

7.1 INTRODUCTION

This section presents a new theoretical approach to multiobject tracking problems which might provide a framework for developing effective analytical tools for multiobject tracking problems. The approach is based on a random spatial point process model of measurements and a Laplace transform method for calculating the likelihood formulas that are essential to deriving tracking algorithms and to analyzing tracking performance. The model includes the one described above, but it is much more compact to express and simpler to understand (once easy, but unfamiliar mathematics is grasped). Moreover, we show how to use simple Laplace transform arguments to analyze this model and avoid some of the notational headache inherent in previous approaches.

Theoretically, our approach is close to and was inspired by the random space-time process work of [22],[23]. However, the space-time processes and the problem described in this section are different from [22],[23]. This section considers only spatial point processes and time is discrete.

In the next subsection we present the mathematical model and background concerning random point processes needed to understand and analyze the model. In subsection 7.3 we state and derive the likelihood ratio formulas for such random point process measurements using a simple Laplace transform approach. Subsections 7.4 and 7.5 illustrate the use of the resulting likelihood ratio formula by deriving the (optimal Bayesian) hypothesis tree type algorithm

(subsection 7.4) and showing how to compute Cramer-Rao lower bounds of mean square estimation error covariances (subsection 7.5). Subsection 7.6 concludes the section.

7.2 RANDOM POINT PROCESS MODELS

Conceptually, the approach of this section is to consider the observations occurring in one time period or scan of data as an image of points rather than as a list of measurements. The approach is intuitively appealing because it corresponds to one's natural idea of radar and sonar displays — devices that provide two-dimensional images of each scan of data. We will show in this section and the next that the approach is also technically appealing because it allows more compact and clearer formulation of models and derivation of estimators. The approach uses the basic ideas of abstract measure theory but only the simplest results. The reader can find the necessary theory in [24],[25].

Following [22],[23], we formulate the image mathematically in terms of a random point process. The precise way to do this is to use random measures [22],[23],[26]. It is mathematically convenient to represent an image of points in terms of a measure taking nonnegative integer values. If μ is such a measure defined with respect to the measure space Y, then $\mu(B)$ is the number of points falling inside a measurable subset B of Y. We will call such integer-valued measures counting measures. An important example of such a measure is the Dirac delta measure denoted δ_y for a given point y of Y and defined so that $\delta_y(B) = 0$ if y is not in B and $\delta_y(B) = 1$ if y is in B. If $y(k)$, $1 \leq k \leq n$, is a sequence of points in Y, then the sum

$$\mu = \sum_{k=1}^{n} \delta_{y(k)} \tag{7-1}$$

is also a counting measure. We will use such sums to represent all the counting measures we need to deal with. Note that the representation of μ by the sequence y(k) is unique except for arbitrary ordering of the sequence of points y(k).

To formulate a statistical measurement model we must define random counting measures. Random counting measures can be defined precisely as random variables taking values in M(Y), the collection of counting measures defined with respect to the measure space Y [26]. In particular, if y is a random variable with respect to Y then δ_y defines a random counting measure with respect to M(Y). One can generate other random counting measures by summing several such Dirac delta measures as in Eq. 7-1.

The most familiar and important example of a random counting measures is the Poisson measure. This is called a Poisson process in [26] because it generalizes the conventional Poisson process defined on the real line. The Poisson measure can be defined generally as follows. Let λ be a finite (non-negative, real-valued) measure with respect to Y. Let N denote a Poisson random variable with mean value equal to $\lambda(Y)$, and let y(k), $1 \leq k$, denote an infinite sequence of independent random variables with respect to Y which are independent of N and which all have the probability distribution given by $\lambda(Y)^{-1}\lambda$. The summation

$$\nu = \sum_{k=1}^{N} \delta_{y(k)} \tag{7-2}$$

defines a Poisson measure with intensity measure λ. The conventional Poisson process N(t) is related by $n(t) = \nu([0,t])$ to the Poisson measure ν whose intensity measure is the Lebesgue measure on the real line.

A Poisson measure has the property that for each measurable subset B of Y, the integer-valued random variable $\nu(B)$ is Poisson distributed with mean value $\lambda(B)$. In particular, one has that

$$E\{\nu(B)\} = \lambda(B) \quad . \tag{7-3}$$

Note that a Poisson measure also has the property that the random variables $\nu(B_1)$ and $\nu(B_2)$ are independent if the sets B_1 and B_2 are disjoint. These properties make the Poisson measures a reasonable model for clutter in multi-object tracking problems. The Poisson distributed random variable $\nu(B)$ models the number of false alarms appearing in the subset B of the observation space in one scan of data. The intensity $\lambda(B)$ is thus the expected number of false alarms appearing in B in one scan.

We can now specify the random point process model we use for multiobject tracking. Let x(t) denote the joint vector of individual target states $x_i(t)$, $1 \le i \le n$,. For simplicity we will assume that there is a known finite number n of targets and that the joint process x(t) is a Markov chain taking values in the measure space X. In subsection 7.4 we will further assume that the chains $x_i(t)$ are independent, but no such assumption is needed in subsection 7.3. The exact nature of these chains does not affect the results obtained here; for example, one can suppose that they are independent, Gaussian processes generated by finite dimensional linear, Gaussian systems.

The measurement process is a sequence μ_t of random counting measures which are conditionally independent given the state process x(t) and whose conditional distribution at time t depends only on x(t). Define μ_t as the sum

$$\mu_t = \tau_t + \nu_t \tag{7-4}$$

where ν_t is a Poisson measure with intensity λ, and τ_t is a random counting measure of the form

$$\tau_t = \sum_{i=1}^{n} z_i(t)\ \delta_{y_i(t)} \quad . \tag{7-5}$$

In Eq. 7-5 the random variable $z_i(t)$ is 1 or 0, depending on whether the i^{th} target is observed (i.e., has a return) in scan t or not. The random variable $y_i(t)$ takes values in Y and is the observation one receives in scan t from the i^{th} target provided that it is detected. We assume that the Poisson measures ν_t are independent of each other and of the processes x(t), $z_i(t)$, and $y_i(t)$ for all i and t. The random variables $z_i(t)$ and $y_i(t)$ are conditionally independent for all i and t given the state process x(t). Furthermore, the conditional distribution of $y_i(t)$ is given by the measure $B \rightarrow H_i\big(B|x(t)\big)$ with respect to the measure space Y, and the conditional probability that $z_i(t) = 1$ is given by $G_i\big(x(t)\big)$. We will make further assumptions concerning these distributions in subsections 7.3 and 7.4.

The model in Eq. 7-4 succinctly represents a scan of data that consists of false alarms (modeled by ν_t) and target detections (modeled by τ_t). The false alarms are modeled appropriately by the Poisson measure ν_t which assumes as a reasonable approximation that false alarms are spatially independent of each other. Note that the intensity measure λ allows us to vary the expected rate of false alarms over the observation space Y. For example, the observation space Y might simply be the plane and λ might be a constant multiple of the uniform distribution over a disk representing the area of coverage of the sensor. This implies a constant rate and density of false alarms over the sensor's area of coverage. The target detections are modeled by the random

counting measure τ_t given in Eq. 7-5. This model assumes that the group of targets produces, at most, n measurements. In subsection 7.4 we will assume additionally that

$$H_i(B|x) = H(B|x_i) \tag{7-6}$$

and

$$G_i(x) = G(x_i) \tag{7-7}$$

so that each target produces, at most, one observation and each observation comes from, at most, one target. The probability distribution $H(B|x_i)$ corresponds to the observation of a detected target; for example, it might be a multivariate Gaussian distribution whose covariance is constant and whose mean is a linear function of $x_i(t)$. The quantity $G(x_i)$ is the detection probability for a target in state x_i; for example, this might be a constant independent of x_i.

In the next section we will derive a likelihood function for the measurement μ_t given the state x(t). Before doing so, we need to present a few more basic mathematical results about random counting measures. Following [26], it is convenient to denote by $\mu(\phi)$ the ordinary Lebesgue-Stieltjes integral of a real-valued, integrable function ϕ with respect to a measure μ. That is, we define

$$\mu(\phi) = \int \phi \, d\mu \quad . \tag{7-8}$$

If μ is a counting measure given by Eq. 7-1, then this integral is just the finite sum

$$\mu(\phi) = \sum_{k=1}^{n} \phi\big(y(k)\big) \quad . \tag{7-9}$$

If the function ϕ is nonnegative, the integral $\mu(\phi)$ is always well-defined, although it may be infinite. If ϕ is a bounded real function and μ is a finite measure, then ϕ is also integrable. If μ is a random counting measure, a stochastic integral can be defined just as in Eq. 7-9. In our examples, all random counting measures are finite so that $\mu(\phi)$ is a well-defined integral for bounded real functions ϕ.

Stochastic integrals such as $\mu(\phi)$ are important because we use them to define the Laplace transform of the random measure μ by

$$L_\mu(\phi) = E\{\exp(-\mu(\phi))\} \tag{7-10}$$

for all nonnegative real function ϕ [26]. The Laplace transform L_μ uniquely determines the distribution of the random measure μ and is often easier to work than the probability distribution of a random measure. An important special case is the Laplace transform of the Poisson measure ν given by

$$L_\nu(\phi) = \exp(-\lambda(1-\exp(-\phi))) \tag{7-11}$$

where λ is the intensity measure of ν and $\lambda(1-\exp(-\phi))$ is the integral of $1 - \exp(-\phi)$ with respect to λ. Note that if λ is a finite measure, then ν is almost surely a finite counting measure and Eq. 7-11 is true for bounded ϕ as well as nonnegative ϕ. As a final example, suppose that τ is the Dirac delta measure δ_y and y is a random variable with probability measure H. Then the Laplace transform of τ is

$$L_\tau(\phi) = H(\exp(-\phi)) \quad . \tag{7-12}$$

In the next subsection we will use the results (Eqs. 7-11 and 7-12) to derive likelihood function for the multiobject tracking measurement model as expressed in Eqs. 7-4 and 7-5.

7.3 LIKELIHOOD RATIO FUNCTIONS

The key to developing an estimation algorithm for the measurement model given in Eqs. 7-4 and 7-5, is to express the likelihood $p(\mu_t|x(t))$ of observing μ_t given the true state is $x(t)$. Since μ_t is a rather complicated mathematical object (namely a counting measure), it is not obvious what the likelihood function $p(\mu|x)$ should be. Precisely defined, the likelihood or likelihood ratio function is a Radon-Nikodym derivative of the observation probability measure with respect to another probability measure, usually of an observation noise [27]. In our case, the other probability measure will be the clutter measure ν_t.

Let $P_1(F|x)$ denote the conditional probability that the observation μ_t lies in the measurable set F of M(Y) given that the true state $x(t) = x$, and let $P_0(F)$ denote the probability that the Poisson measure ν_t lies in F. We are looking for the Radon-Nikodym derivative of P_1 with respect to P_0. That is, the likelihood function $p(\mu|x)$ is a real-valued function of the set M(Y) of counting measures μ and the set X of states x which satisfies the relation

$$P_1(F|x) = \int_F p(\xi|x)\, dP_0(\xi) \quad . \tag{7-13}$$

Because the distribution P_1 is uniquely determined by its Laplace transform [26], proving Eq. 7-13 is equivalent to proving

$$\int \exp(-\xi(\phi))\, dP_1(\xi) = \int \exp(-\xi(\phi))\, p(\xi|x)\, dP_0(\xi) \tag{7-14}$$

for all nonnegative, measurable ϕ. Written in terms of expectations, Eq. 7-14 is equivalent to

$$E\{\exp(-\mu_t(\phi))\,|\,x(t)\} = E\{p(\nu_t|x(t))\exp(-\nu_t(\phi))\,|(t)\} \quad . \tag{7-15}$$

We can rewrite the left side of Eq. 7-15 using Eqs. 7-4, 7-5 and the Laplace transform relations (Eqs. 7-11 and 7-12). Note also that the random measure ν_t is independent of x(t), and the random variables $z_i(t)$ and $y_i(t)$ are conditionally independent of each other and ν_t given x(t). Thus, we have

$$E\{\exp(-\mu_t(\phi))|x(t)\} = \prod_{k=1}^{n} [1-G_i + G_iH_i(\exp(-\phi))] \exp(-\lambda(1-\exp(-\phi))) \quad . \tag{7-16}$$

In Eq. 7-16 we have suppressed the dependence on x(t) of the conditional detection probabilities G_i and the conditional observation probabilities H_i. We will continue to do this to simplify our notation.

Our objective is to find a function $p(\mu|x)$ such that

$$\prod_{k=1}^{n} [1-G_i + G_iH_i(\exp(-\phi))] \exp(-\lambda(1-\exp(-\phi))) = E\{p(\nu_t|x(t))\exp(-\nu_t(\phi))|x(t)\} \tag{7-17}$$

We will do this in a series of increasingly more complicated results, beginning with the case of a single target with perfect detection (i.e., for which n = 1 and $G_i(x) = 1$ for all x). In this simple case, the basic ideas are clearest. The treatment of imperfect detection is a simple corollary. Next follows the case of n targets with perfect detection and finally the general case.

Proposition 1

Assume that the conditional probability H_1 has a density k_1 with respect to the intensity measure λ; that is, assume that

$$H_1(B|x) = \int k_1(y|x)\, d\lambda(y) \quad . \tag{7-18}$$

Then for n = 1 and $G_1 = 1$, $p(\mu|x)$ is given by

$$p(\mu|x) = \mu(k_1) \quad . \tag{7-19}$$

Note that Eq. 7-19 means that

$$p(\mu|x) = \int k_1(y|x)\, d\mu(y) = \sum_{j=1}^{n} k_1(y(j)|x) \tag{7-20}$$

if the counting measure μ is given by the sum in Eq. 7-1.

The assumption about the measurement density in this proposition is basically that the conditional observation probability distribution H_1 is absolutely continuous [24],[25] with respect to the clutter measure λ. In more physical terms, we are assuming that no measurements will occur where there are not some false alarms. In many cases, the measures H_1 and λ will have densities h_1 and η with respect to Lebesgue measure defined on the observation space $Y = R^m$. For example, this is the case if one's observations have continuous values and false alarms occur uniformly over the area of sensor coverage. Then k_1 is given by the ratio

$$k_1(y|x) = \frac{h_1(y|x)}{\eta(y)} \quad , \tag{7-21}$$

provided that $h_1(y|x) = 0$ whenever $\eta(y) = 0$ so that the ratio is well-defined (this condition is the substance of the assumption in Proposition 1).

Proof of Proposition

Suppose that $\phi(y)$ and $k_1(y|x)$ are bounded as functions of y for each x and define

$$\phi_\varepsilon(y) = \phi(y) - \varepsilon\, k_1(y|x(t)) \tag{7-22}$$

for all real ε. From Eq. 7-11 and the independence of x(t) and ν_t note that

$$\exp(-\lambda(1-\exp(-\phi_\varepsilon))) = E\{\exp(-\nu_t(\phi_\varepsilon))|x(t)\} \quad . \tag{7-23}$$

Under the boundedness conditions we can differentiate Eq. 7-23 with respect to ε. Doing so at $\varepsilon = 0$ gives

$$\lambda(k_1 \exp(-\phi)) \exp(-\lambda(1-\exp(-\phi))) = E\{\nu_t(k_1 \exp(-\nu_t(\phi)))|x(t)\} \qquad (7\text{-}24)$$

which gives the desired result upon noting $\lambda(k_1 \exp(-\phi)) = H_1(\exp(-\phi))$. The result for unbounded, nonnegative ϕ and k_1 follows from standard application of Legesgue's dominated and monotone convergence theorems [24],[25].

Corollary 1

Assume the same hypothesis as in Proposition 1 but do not assume that $G_1(x) = 1$. Then $p(\mu|x)$ is given by

$$p(\mu|x) = 1 - G_1(x) + G_1(x)\ \mu(k_1) \qquad (7\text{-}25)$$

Proof of Corollary 1

Note that from Eq. 7-24 in Proposition 1 we have

$$G_1 H_1(\exp(-\phi)) \exp(-\lambda(1-\exp(-\phi))) = E\{G_1 \nu_t(k_1)\exp(-\nu_t(\phi))|x(t)\} \qquad (7\text{-}26)$$

From the basic formula (Eq. 7-23) we have

$$(1-G_1) \exp(-\lambda(1-\exp(-\phi))) = E\{(1-G_1)\exp(-\nu_t(\phi))|x(t)\} \ . \qquad (7\text{-}27)$$

Adding Eqs. 7-26 and 7-27 together gives the desired result (Eq. 7-17) for the case n=1.

The result for multiple targets is more complicated and requires some preliminary notation and definitions. For any counting measure μ on Y define the new counting measure $\mu^{(n)}$ on Y^n as follows. For each bounded real function f on Y^n define the integral

$$\mu^{(n)}(f) = \sum \alpha_n(i_1,\ldots,i_n)\ f(y(i_1),\ldots,y(i_n)) \qquad (7\text{-}28)$$

where the sum is over all sequences $i_1,\ldots,i_n$ of integers such that $1 \le i_j \le n$, and $\alpha_n = 0$ if $i_j = i_k$ for $j \neq k$, $\alpha_n = 1$ otherwise. Note that $\alpha_1(y(1)) = 1$ for all $y(1)$. The sequence $y(k)$ is the one used to represent μ in Eq. 7-1. It should be clear that the definition does not depend on the order of the $y(k)$. The sum in Eq. 7-28 defines the integral of a counting measure on Y^n. Note that the expression in Eq. 7-28 can be interpreted as the sum over all sequences $i_1,\ldots,i_n$ such that no i_j appears twice in the same sequence. Also note that $\mu^{(1)} = \mu$.

Finally, if $f_1,\ldots,f_n$ are function of Y, define the product by

$$f_1 \times \ldots \times f_n(y_1,\ldots,y_n) = f_1(y_1)\ldots f_n(y_n) \quad . \qquad (7\text{-}29)$$

We are now ready to state and prove the results for multiple targets, first for the perfect detection case.

Proposition 2

Assume that the conditional probability H_i has a density k_i with respect to the intensity measure λ for each $i=1,\ldots,n$. Assume also that the detection probability $G_i(x) = 1$ for all i and x. Then the likelihood function $p(\mu|x)$ is given by

$$p(\mu|x) = \mu^{(n)}(k_1 \times \ldots \times k_n) \quad . \qquad (7\text{-}30)$$

Note that Eq. 7-30 means that

$$\begin{aligned} p(\mu|x) &= \int \prod_{j=1}^{n} k_j(y_j|x)\ d\mu^{(n)}(y_1,\ldots,y_n) \\ &= \sum \alpha_n(i_1,\ldots,i_n) \prod_{j=1}^{n} k_j(y(i_j)|x) \quad . \end{aligned} \qquad (7\text{-}31)$$

Proof of Proposition 2

We need to prove Eq. 7-17 is true when $G_i = 1$ and $p(\mu|x)$ is given by Eq. 7-30. As in the proof of Proposition 1, it suffices to prove this for the case when ϕ and k_i are bounded functions with respect to y. We have proven the case n = 1 in Proposition 1. We will prove the general case by induction on n. Thus, assume that Proposition 2 is true for n, and let us deduce the result for n + 1.

The result for n states that

$$\prod_{i=1}^{n} \left[H_i(\exp(-\phi))\right] \exp\left(-\lambda(1-\exp(-\phi))\right) = E\left\{\nu_t^{(n)}(k_1 \times \ldots \times k_n)\exp\left(-\nu_t(\phi)\right)|x(t)\right\} \tag{7-32}$$

As in Proposition 1, substitute the perturbation ϕ_ε

$$\phi_\varepsilon = \phi - \varepsilon\, k_{n+1} \tag{7-33}$$

for ϕ in Eq. 7-32 and take the derivative with respect to ε. This gives the following result when $\varepsilon = 0$:

$$\prod_{i=1}^{n} \left[H_i \exp(-\phi)\right] \exp\left(-\lambda(1-\exp(-\phi))\right)$$

$$= E\left\{\nu_t^{(n)}(k_1 \times \ldots \times k_n)\ \nu_t(k_{n+1})\ \exp\left(-\nu_t(\phi)|x(t)\right)\right\}$$

$$- \sum_{j=1}^{n} \lambda\left(k_j k_{n+1} \exp(-\phi)\right) \prod_{i \neq j}^{n} \left[H_i(\exp(-\phi))\right] \exp\left(-\lambda(1-\exp(-\phi))\right) . \tag{7-34}$$

Substituting $k_j k_{n+1}$ for k_j in the result (Eq. 7-32) for n gives

$$\lambda\left(k_j k_{n+1} \exp(-\phi)\right) \prod_{i \neq j}^{n} \left[H_i(\exp(-\phi))\right] \exp\left(-\lambda(1-\exp(-\phi))\right)$$

$$= E\left\{\nu_t^{(n)}(k_1 \times \ldots \times k_j k_{n+1} \times \ldots \times k_n)\ \exp\left(-\nu_t(\phi)\right)|x(t)\right\} . \tag{7-35}$$

Finally, note that for any counting measure μ and functions k_i defined on Y defined on Y we have that

$$\mu^{(n+1)}(k_1 \times \ldots \times k_{n+1}) = \mu^{(n)}(k_1 \times \ldots \times k_n)\mu(k_{n+1}) - \sum_{j=1}^{n} \mu^{(n)}(k_1 \times \ldots \times k_j k_{n+1} \times \ldots \times k_n) \quad . \tag{7-36}$$

Putting Eqs. 7-35 and 7-36 in Eq. 7-34 gives the desired result for n + 1.

The general case with imperfect detection follows Proposition 2 in the same way that Corollary 1 followed from Proposition 1. The notation, however, is more complicated. Adjoin a new point θ to the observation set Y and call the extended set $\underline{Y}$. Extend any counting measure μ defined on Y to a counting measure $\underline{\mu}$ defined on $\underline{Y}$ by defining $\underline{\mu}(B)$ if θ is not in B and $\underline{\mu}(B) + 1$ if θ is in B. The point θ represents a fictitious measurement indicating a possible missed detection and is always counted by $\underline{\mu}$. Likewise, extend $\mu^{(n)}$ to $\underline{\mu}^{(n)}$ as follows. Define $\underline{\alpha}_n$ as α_n was defined before except that $\underline{\alpha}_n = 0$ if $i_j = i_k$ for $j \neq k$ and if $y(i_j) \neq \theta$. Note that $\underline{\mu}^{(n)}$ can be interpreted similarly to $\mu^{(n)}$. The sum corresponding to Eq. 7-28 is over all sequences $i_1, \ldots, i_n$ such that no i_j appears twice in the same sequence, except that the index i for which $y(i) = \theta$ can occur any number of times.

It is also necessary to extend k_i to a function $\underline{k}_i$ defined on $\underline{Y}$ as follows:

$$\underline{k}_i(y|x) = G_i(x)\ k_i(y|x) \tag{7-37}$$

for y in Y, and

$$\underline{k}_i(\theta|x) = 1 - G_i(x) \quad . \tag{7-38}$$

Corollary 2

Assume the same hypothesis as in Proposition 2 but do not assume that $G_i(x) = 1$. Then $p(\mu|x)$ is given by

$$p(\mu|x) = \underline{\mu}^{(n)}(\underline{k}_1 \times \ldots \times \underline{k}_n) \quad . \tag{7-39}$$

Proof of Corollary 2

The proof follows that of Proposition 2 with minor variations. We have proven the n=1 case in Corollary 1. To prove the general case by induction, assume that the following result is true for n and all bounded functions ϕ, k_j and detection probability functions G_j.

$$\begin{aligned} &\prod_{i=1}^{n} \left[1-G_i + G_i H_i\left(\exp(-\phi)\right)\right] \exp\left(-\lambda(1-\exp(-\phi))\right) \\ &\quad = E\left\{\underline{\nu}_t^{(n)}(\underline{k}_1 \times \ldots \times \underline{k}_n) \exp\left(-\nu_t(\phi)\right) | x(t)\right\} \quad . \end{aligned} \tag{7-40}$$

Substitute the perturbation

$$\phi_\varepsilon = \phi - \varepsilon G_{n+1} k_{n+1} \tag{7-41}$$

for ϕ in Eq. 7-40. The derivative at $\varepsilon=0$ gives

$$\begin{aligned} &G_{n+1} H_{n+1}\left(\exp(-\phi)\right) \prod_{i=1}^{n} \left[1-G_i + G_i H_i\left(\exp(-\phi)\right)\right] \exp\left(-\lambda(1-\exp(-\phi))\right) \\ &\quad = E\left\{\underline{\nu}_t^{(n)}(\underline{k}_1 \times \ldots \times \underline{k}_n)\ G_{n+1}\nu(k_{n+1}) \exp\left(-\nu_t(\phi)\right) | x(t)\right\} \\ &\quad - \sum_{j=1}^{n} G_{n+1} G_j H_j\left(k_{n+1}\exp(-\phi)\right) \prod_{i\neq j}^{n} \left[1-G_i+G_i H_i\left(\exp(-\phi)\right)\right] \\ &\qquad \exp\left(-\lambda(1-\exp(-\phi))\right) \quad . \end{aligned} \tag{7-42}$$

Suppose that in Eq. 7-40 we substitute the function $\beta_j = G_{n+1} G_j k_{n+1} k_j$ for k_j and 1 for G_j. Then we obtain

$$E\{\underline{\nu}_t^{(n)}(\underline{k}_1\times\ldots\times\underline{\beta}_j\times\ldots\times\underline{k}_n)\ \exp(-\nu_t(\phi))|x(t)\}$$

$$= G_{n+1}G_jH_j(k_{n+1}\exp(-\phi))\prod_{i\neq j}^{n}[1-G_i + G_iH_i(\exp(-\phi))]\exp(-\lambda(1-\exp(-\phi)))\ . \tag{7-43}$$

Multiply Eq. 7-40 by $1 - G_{n+1}$ to get

$$(1-G_{n+1})\prod_{i=1}^{n}[1-G_i + G_iH_i(\exp(-\phi))]\exp(-\lambda(1-\exp(-\phi))) = E(\underline{\nu}_t^{(n)}(\underline{k}_1\times\ldots\times\underline{k}_n)(1-G_{n+1})\ \exp(-\nu_t(\phi))|x(t)\}\ . \tag{7-44}$$

Adding Eq. 7-44 to Eq. 7-42 and using the relation 7-43 gives

$$\prod_{i=1}^{n+1}[1-G_i + G_iH_i(\exp(-\phi))]\exp(-\lambda(1-\exp(-\phi))) = E\{[\underline{\nu}_t^{(n)}(\underline{k}_1\times\ldots\times\underline{k}_n)\underline{\nu}_t(\underline{k}_{n+1}) \tag{7-45}$$

$$-\sum_{j=1}^{n}\underline{\nu}_t^{(n)}(\underline{k}_1\times\ldots\times\underline{\beta}_j\times\ldots\times\underline{k}_n)]\exp(-\nu_t(\phi))|x(t)\}\ . \tag{7-45}$$

Finally, let us show that the right side of Eq. 7-45 is what we want. Corresponding to Eq. 7-35 is a recursive expression for $\underline{\alpha}_n$ given by

$$\underline{\alpha}_{n+1}(i_1,\ldots,i_{n+1}) = \underline{\alpha}_n(i_1,\ldots,i_n)\left[1 - \sum_{j=1}^{n}\underline{\delta}(i_j,i_{n+1})\right] \tag{7-46}$$

where $\underline{\delta}(i,j) = 1$ only if $i=j$ and $y(i) \neq \theta$ ($\underline{\delta}$ is 0 otherwise). This recursion implies that for any counting measure μ defined on Y and functions $\underline{k}_j$ defined on the extended set $\underline{Y}$ we have

$$\underline{\mu}^{(n+1)}(\underline{k}_1\times\ldots\times\underline{k}_{n+1}) = \underline{\mu}^{(n)}(\underline{k}_1\times\ldots\times\underline{k}_n)\mu(\underline{k}_{n+1}) - \sum_{j=1}^{n}\underline{\mu}^{(n)}(\underline{k}_1\times\ldots\times\underline{\beta}_j\times\ldots\times\underline{k}_n) \tag{7-47}$$

where $\underline{\beta}_j$ is defined so that $\underline{\beta}_j(y) = \underline{k}_j(y)\underline{k}_{n+1}(y)$ for y in Y and $\beta_j(\theta) = 0$. If $\underline{k}_j$ and $\underline{k}_{n+1}$ are defined as in Eqs. 7-37 and 7-38, then $\underline{\beta}_j$ in Eq. 3-37 is precisely the extension of β_j in Eq. 7-43 to $\underline{Y}$. Thus, substituting Eq. 7-47 with $\mu = \nu_t$ into Eq. 7-45 gives the desired result.

7.4 ESTIMATION ALGORITHM

The conditional probability distribution of the state variable x(t), given all observations up to time period t, contains all the statistical information about x(t) that exists in the measurements. Let $\pi_t(x)$ denote the density of the conditional probability distribution with respect to some measure γ (γ might be Lebesgue measure, for example). Similarly, let $\pi_{t+1,t}(x)$ denote the conditional probability density of the state x(t+1) given all observations up to the time t. The recursive procedure for computing $\pi_{t+1}(x)$ from a new observation and the predicted density $\pi_{t+1,t}(x)$ is Bayes' rule. In terms of a likelihood function $p(\mu_t|x)$ for the measurement model, Bayes' rule is

$$\pi_{t+1}(x) = \frac{\pi_{t+1,t}(x)\ p(\mu_t|x)}{\int \pi_{t+1,t}(\xi)\ p(\mu_t|\xi)\ d\gamma(\xi)} \quad . \tag{7-48}$$

In this section we sketch briefly how Bayes' rule applies to the likelihood function of Eq. 7-39 in the previous subsection leads directly to the measurement update computation found in hypothesis tree type multiobject tracking algorithms.

Hypothesis tree type estimation algorithms are based on the representation of the conditional probability density of the state as a weighted sum of simpler densities, namely,

$$\pi_t(x) = \sum_j \beta_j \, \pi_j(x) \tag{7-49}$$

where each $\pi_j(x)$ is a product

$$\pi_j(x) = \prod_{i=1}^{n} \pi_{i,j}(x_i) \quad . \tag{7-50}$$

The density $\pi_{i,j}(x_i)$ represents the conditional probability of the i^{th} target being in state x_i given that hypothesis j is true. The weight β_j is the probability that hypothesis j is true. For example, in applications, the densities $\pi_{i,j}$ and π_j might be approximated by Gaussian distributions so that Eq. 7-49 is a Gaussian sum distribution.

Using the likelihood function of subsection 7.3, we will derive the expression (Eq. 7-49) for $\pi_{t+1}(x)$. First assume that x(t) is the joint vector of individual target states $x_i(t)$, $1 \leq i \leq n$. For simplicity we will assume that there is a known finite number n of targets and that the processes $x_i(t)$ are independent Markov chains. Assume that the measurement μ_t is modeled as in subsection 7.2 and that the measurement densitites $h_i(y|x)$ satisfy

$$h_i(y|x) = h(y|x_i) \quad , \tag{7-51}$$

and the detection probabilities $G_i(x)$ satisfy

$$G_i(x) = G(x_i) \quad . \tag{7-52}$$

Let $\eta(y)$ denote the false alarm density, and as in subsection 7.3, assume that $\eta(y)$ is not 0 unless $h(y|x_i)$ is 0 for all values of x_i. In this case, $\underline{k}_i$ is given by

$$\underline{k}_i(y|x) = \underline{k}(y|x_i) = \frac{G(x_i)\ h(y|x_i)}{\eta(y)} \tag{7-53}$$

for y in the observation space Y (i.e., y corresponding to a detection), and is given by

$$\underline{k}_i(y|x) = \underline{k}(y|x_i) = 1 - G(x_i) \tag{7-54}$$

For $y = \theta$ (i.e., for y corresponding to a missed detection). The algorithm for updating Eq. 7-49 is then given in terms of a prediction step and a measurement step as follows.

Prediction Step

For each hypothesis index j and each target index i, predict ahead the density $\pi_{i,j}$ one time period to obtain the conditional density $\pi_{i,j}^{+}(x_i)$ of $x_i(t+1)$ being in state x_i given measurements up to time period t. The predicted density is given by the weighted sum of products

$$\pi_{t+1,t}(x) = \sum_j \beta_j \prod_{i=1}^{n} \pi_{i,j}^{+}(x_i) \quad . \tag{7-55}$$

For linear Gaussian target models, this step involves just the prediction of the conditional mean and covariance. For nonlinear target models, this step would require propagation of some other statistics describing the conditional density or an approximation (such as in an extended Kalman filter). Note that the prediction step does not change the hypotheses j or the probabilities β_j of the hypotheses. This occurs in the next step when measurements at time period t+1 are processed.

To work out the measurement processing step, one substitutes the expression 7-55 for $\pi_{t+1,t}(x)$ and the likelihood function 7-39 for $p(\mu_{t+1}|x)$ into Bayes' rule (Eq. 7-48). Let us carry this out in detail. Suppose that the point process measurement μ_{t+1} consists of the observations y(k), $1 \leq k \leq m$,

listed in some arbitrary order (compare Eq. 7-1). Let y(m+1) denote the fictitious observation y(m+1) = θ. Then the likelihood function $p(\mu_{t+1}|x)$ of Eq. 7-39 is given by the summation

$$\begin{aligned} p(u_{t+1}|x) &= \underline{\mu}_{t+1}{}^{(n)}(\underline{k}_1 x \ldots \underline{k}_n) \\ &= \sum \underline{\alpha}_n(i_1,\ldots,i_n)\ \underline{k}(y(i_1)|x_1)\ldots\underline{k}(y(i_n)|x_n) \end{aligned} \tag{7-56}$$

where the sum is over all sequence $i_1,\ldots,i_n$ of integers such that $1 \le i_j \le m+1$, and $\underline{\alpha}_n = 0$ only if $i_j = i_k$ for $j \neq k$ and $i_j \neq m+1$, else $\underline{\alpha}_n = 1$. Note that a sequence $i_1,\ldots,i_n$ associates target j with one of the m observations $y(i_j)$ or with the fictitious measurement θ (to denote the lack of real observation for that target). The coefficient $\underline{\alpha}_n$ indicates which associations are valid, namely, those which associate each real measurement with, at most, one target.

Before presenting the update algorithm, it is necessary to introduce some further notation. If measurement k is associated with target i of hypothesis j, then the conditional density $\pi_{i,j}{}^{+}$ is updated to $\pi_{i,j,k}$ by

$$\pi_{i,j,k}(x_i) = \frac{\pi_{i,j}{}^{+}(x_i)G(x_i)h(y(k)|x_i)}{\int \pi_{i,j}{}^{+}(\xi_i)G(\xi_i)h(y(k)|\xi_i)\ d\gamma(\xi_i)} \tag{7-57}$$

if $1 \le k \le m$, and

$$\pi_{i,j,m+1}(x_i) = \frac{\pi_{i,j}{}^{+}(x_i)[1-G(x_i)]}{\int \pi_{i,j}{}^{+}(\xi_i)[1-G(\xi_i)]\ d\gamma(\xi)} \tag{7-58}$$

for k = m+1 (processing the missed detection). Similarly, the likelihood that the measurement k is associated with target i of hypothesis j is given by

$$\sigma_{i,j,k} = \int \frac{\pi_{i,j}{}^{+}(\xi_i)G(\xi_i)h(y(k)|\xi_i)}{\eta(y(k))}\ d\gamma(\xi_i) \tag{7-59}$$

if $1<k<m$, and by

$$\sigma_{i,j,m+1} = \int \pi_{i,j}^{+}(\xi_i)[1-G(\xi_i)]\, d\gamma(\xi_i) \tag{7-60}$$

for $k = m+1$ in the missed detection case.

Using this notation, substitute Eqs. 7-56 and 7-57 to obtain

$$\pi_{t+1,t}(x)\; p(\mu_t|x) = \sum_j \sum_i \beta_j\, \underline{\alpha}_n(i_1,\ldots,i_n) \prod_{k=1}^{n} \sigma_{k,j,i_k} \pi_{k,j,i_k}(x_k) \quad . \tag{7-61}$$

Thus, Bayes' rule (Eq. 7-48) gives

$$\pi_{t+1}(x) = \frac{\sum_j \sum_i \beta_j\, \underline{\alpha}_n(i_1,\ldots,i_n) \prod_{k=1}^{n} \sigma_{k,j,i_k} \pi_{k,j,i_k}(x_k)}{\sum_j \sum_i \beta_j\, \underline{\alpha}_n(i_1,\ldots,i_n) \prod_{k=1}^{n} \sigma_{k,j,i_k}} \tag{7-62}$$

The update step of the estimation algorithm is then given as follows.

Update Step

For each hypothesis j, each target index i and each measurement index k, update the density $\pi_{i,j}^{+}$ to obtain the updated target densities $\pi_{i,j,k}(x_i)$ of $x_i(t+1)$ being in state x_i given hypothesis j was true up to time t and that measurement k was associated with target i at time t+1. Similarly, calculate the likelihood $\sigma_{i,j,k}$ of measurement k being associated with target i given hypothesis j was true up to time t. For each hypothesis j at time t and for each association $i_1,\ldots,i_n$ of measurements with targets, generate a new hypothesis j' at time t+1 with probability

$$\beta_{j'} = \frac{\beta_j\, \underline{\alpha}_n(i_1,\ldots,i_n) \prod_{k=1}^{n} \sigma_{k,j,i_k}}{\sum_j \sum_i \beta_j\, \underline{\alpha}_n(i_1,\ldots,i_n) \prod_{k=1}^{n} \sigma_{k,j,i_k}} \tag{7-63}$$

and with target conditional densitities

$$\pi_{k,j'}(x_k) = \pi_{k,j,i_k}(x_k) \quad . \tag{7-64}$$

The conditional density of the combined target states at time t+1 is then

$$\pi_{t+1}(x) = \sum_{j'} \beta_{j'} \prod_{i=1}^{n} \pi_{i,j'}(x_i) \quad . \tag{7-65}$$

Note the tree structure of the hypotheses. At every update step each hypothesis j generates many new disjoint hypotheses j'. Over time, the hypotheses can be arranged in a tree with the nodes at one level of the tree representing hypotheses at one period of time. Branching of nodes at one level of the tree to nodes at the next level of the tree represents the generation of new hypotheses. Note also that the number of nodes in this tree increases roughly as n^{mt} if m is the number of observations per time period, t is the number of time periods, and n is the (fixed) number of targets. Several approximations have been developed to limit this geometric growth of hypotheses (see [1],[2]).

7.5 APPLICATION TO PERFORMANCE ANALYSIS

In this section we discuss briefly how the result of subsection 7.3 can be used to obtain Cramer-Rao type lower bounds of mean square terror in estimating x(t) given a measurement model of the form described in Eq. 7-4 and 7-5. From Section 4, note that the Cramer-Rao bound depends on the Fisher information defined by

$$I(t) = E\left\{ \frac{1}{p(\mu_t | x(t))^2} \frac{\partial p^T}{\partial x} \frac{\partial p}{\partial x} \right\} \quad . \tag{7-66}$$

The derivatives are computed from

$$\frac{\partial p}{\partial x}(\mu|x) = \sum_{j=1}^{n} \underline{\mu}^{(n)}\left(\underline{k}_1 \times \ldots \times \frac{\partial \underline{k}_j}{\partial x} \times \ldots \times \underline{k}_n\right) . \tag{7-67}$$

It is assumed that the measurement density $k_1(y|x)$ and the detection probability $G_1(x)$ are both differentiable with respect to x. The expressions needed to compute I(t) reduce to

$$p(\mu|x) = 1-G_1(x) + G_1(x)\mu(k_1(y|x)) \tag{7-68}$$

and

$$\frac{\partial p}{\partial x}(\mu|x) = -\frac{\partial G}{\partial x}1(x) + \frac{\partial G}{\partial x}1(x)\mu(k_1(y|x)) + G_1(x)\mu\left(\frac{\partial k}{\partial x}1(y|x)\right) \tag{7-69}$$

for the case of one target (n=1).

Actually computing I(t) requires taking the expectation in Eq. 7-66 with respect to the random variable x(t) and the random measure μ_t. In general, this expectation is not expressible in closed form and it is necessary to resort to analytical (e.g., perturbation expansion) or numerical approximation (e.g., perturbation expansion) or numerical approximation (e.g., Monte-Carlo simulation). Note that once I(t) has been obtained, it is easy to find the Cramer-Rao lower bound on the error covariance for estimating a Gaussian distributed state. Suppose that x(t) satisfies the linear Gaussian equation

$$x(t+1) = A(t)x(t) + B(t)w(t) \tag{7-70}$$

where w(t) is Gaussian white noise with covariance Q(t) and x(1) has prior covariance Σ_0. Initialize P_0 and S_0 to Σ_0, and compute P_t and S_t using the recursion defined by

$$P_{t+1} = [S_t^{-1} + I(t+1)]^{-1} \tag{7-71}$$

for $t>0$ and

$$S_t = A(t)P_tA(t)^T + B(t)Q(t)B(t)^T \quad . \tag{7-72}$$

for $t>1$. If $\hat{x}(t)$ is the minimum mean square estimate of $x(t)$ based on the measurements μ_s for $1<s<t$, then P_t is the Cramer-Rao lower bound on the error covariance

$$P_t \leq E\{[x(t)-\hat{x}(t)][x(t)-\hat{x}(t)]^T\} \quad . \tag{7-73}$$

7.6 CONCLUDING REMARKS

In this section we have introduced a new formulation of multiobject tracking as an estimation problem with random point process measurements and derived the corresponding likelihood ratio formula using Laplace transform methods. This new theoretical approach provides a compact formulation and powerful technique for analyzing multiobject tracking problems. It allows easy derivation of optimal tracking algorithms and may prove useful in analyzing the optimal performance of such algorithms. The approach also extends easily to other kinds of multiobject tracking problems than the one described in subsections 7.3 and 7.4, such as measurement models in which several returns are received from the same target (e.g., multipath phenomena). Although much work remains to be done, we believe that the approach presented in this section provides an effective framework for developing new analytical tools to understand better the quantitative performance of multiobject tracking algorithms.

SECTION 8

NUMERICAL EXAMPLES AND CONCLUSIONS

8.1 NUMERICAL COMPARISON OF PERFORMANCE METHODS

To understand better the relative accuracy of the different performance methods presented in Section 4 through 6, we considered the following simple Type I hybrid state estimation problem:

$$x(t+1) = x(t) + w(t) \tag{8-1}$$

$$y(t) = x(t) + q(t)v_1(t) + \big(1-q(t)\big)v_0(t) \tag{8-2}$$

This is a dynamic version of the simple static problem analyzed in Section 3. We assume that the variance σ_1^2 of $v_1(t)$ is much larger than the variance σ_0^2 of $v_0(t)$. That is, $q(t) = 1$ indicates a bad measurement and $q(t) = 0$ indicates a good measurement.

We describe two cases whose parameters are shown in Table 8-1. The two cases differ only by the initial variance of x(0). The initial variance is small in Case 1 to let us study steady-state estimation performance -- performance achieved once the initial uncertainty in the state estimate has been reduced. The large initial variance in Case 2 lets us study acquisition performance -- the details of how a very large initial uncertainty is reduced until steady-state performance is achieved.

TABLE 8-1. STATISTICAL PARAMETERS FOR NUMERICAL EXAMPLES

	CASE 1	CASE 2
Var $(x(0))$	4.0	10^4
Var (w)	4.0	4.0
Prob $(q=1)$	.1	.1
Var (v_1)	10^4	10^4
Var (v_0)	4.0	4.0

Figures 8-1 and 8-2 show performance estimates for the two cases. In each figure we have plotted four types of performance: Cramer-Rao lower bound (CRB), an upper bound given by the best linear estimator (LIN), simulated performance of a simple gating algorithm (GATE), and simulated performance of the optimal algorithm (OPT). We also computed the rate distortion lower bound of Section 5 for this problem, but the performance prediction was several orders of magnitude worse than the other methods, so we omitted the rate distortion results.

Figure 8-1 shows steady-state performance. The best linear estimator's performance increases since it assumes measurement noise is Gaussian with variance equal to the average variance of $v_1(t)$ and $v_0(t)$,

$$\varepsilon\,\sigma_1^2 + (v-\varepsilon)\sigma_0^2 = (.1)10^4 + (.9)4$$
$$\cong 10^3 \quad .$$

As $t \to \infty$, the best linear estimate will achieve steady-state performance, but only for t much greater than the seven time steps considered in Fig. 8-1.

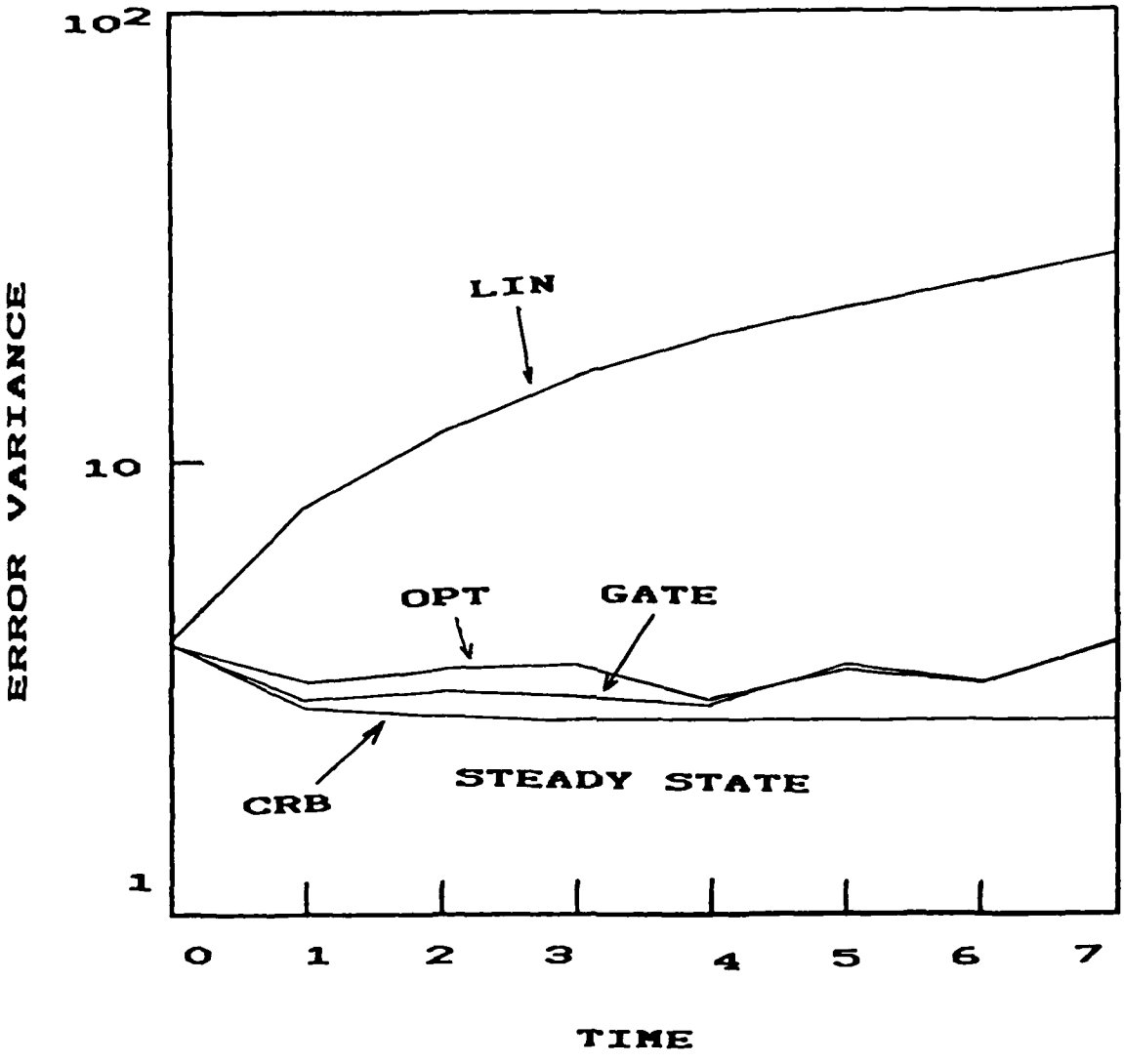


Figure 8-1. Case 1: Steady-State Performance

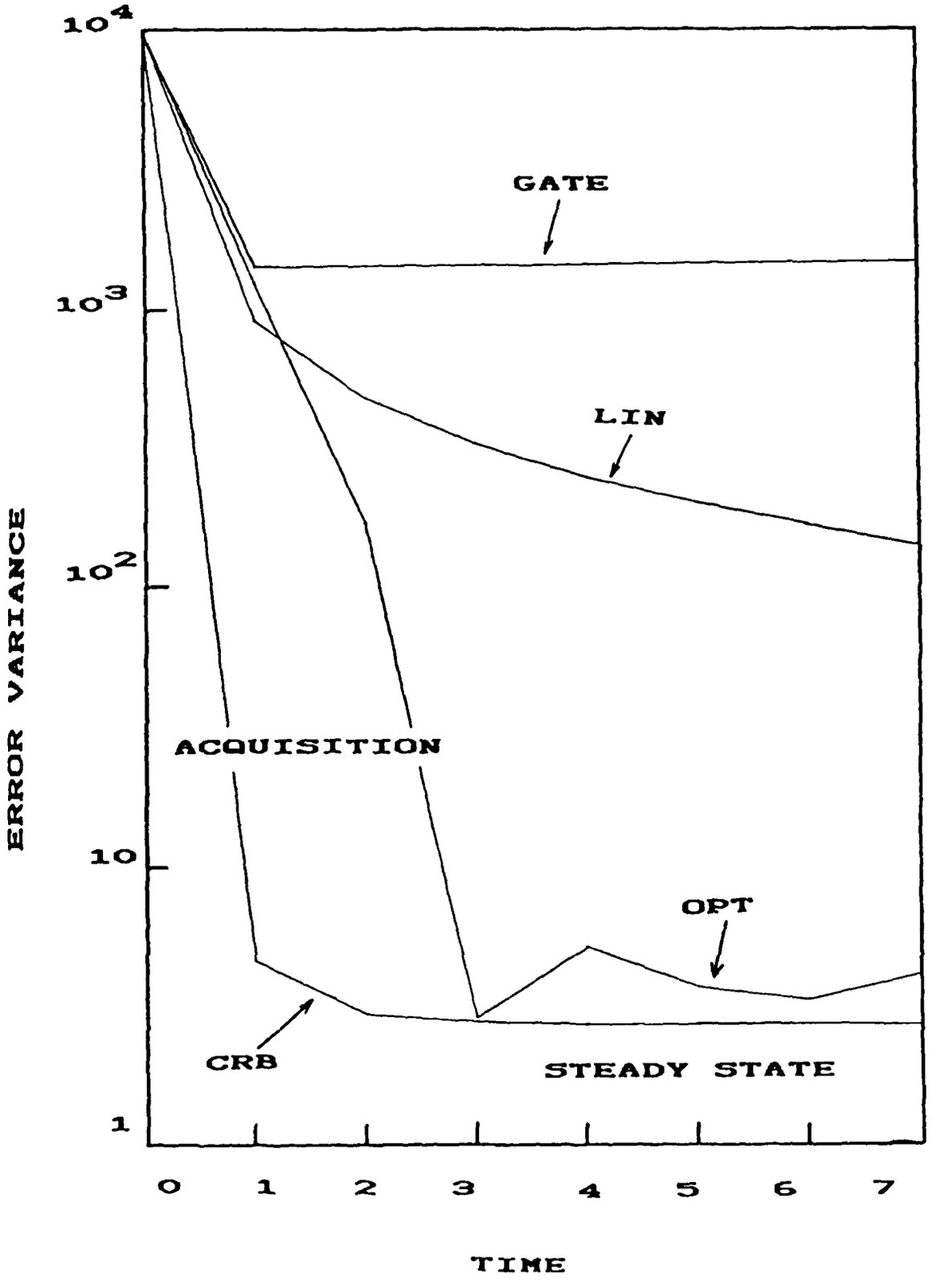


Figure 8-2. Case 2: Acquisition Performance

The Cramer-Rao lower bound, the simulated gate algorithm, and the simulated optimal algorithm all track fairly well in Fig. 8-1. Note that one expects the gate algorithm to be near-optimal in steady-state: that is, it performs well once a good state estimate has been achieved.

Figure 8-2 shows performance when the initial state variance is the same magnitude as bad measurements. Expressed in tracking terms, an estimator must first acquire the target: that is, reduce the initially large error variance to steady-state performance. As shown in Fig. 8-2, the optimal estimator acquires the state after about three measurements. On the other hand, the gate algorithm never acquires the target and the best linear estimator does so only very slowly. The Cramer-Rao lower bound optimistically predicts a quicker acquisition (after 1 measurement), but predicts steady-state performance accurately. Note that the poor performance of the gate algorithm is to be expected. Gate algorithms are designed based on steady-state assumptions and usually require an auxiliary initialization algorithm to acquire the steady-state performance.

As a final numerical example, Table 8-2 shows results comparing the Monte-Carlo method of Section 7 with the simulated optimal algorithm for Case 2 above over the first five measurement updates. We have considered three different choices of the Monte-Carlo parameters n and m (see Section 7). Note that the Monte-Carlo performance method predicts the acquisition phenomena but tends to be pessimistic about steady-state performance.

TABLE 8-2. MONTE-CARLO PERFORMANCE METHOD

	TIME	OPTIMAL MSE	MONTE-CARLO METHOD MSE		
n			100	100	1000
m			100	1000	10
	1	1219	1335	968	3425
	2	171	566	292	1989
	3	2.9	347	65	1395
	4	5.1	184	35	1087
	5	3.7	139	26	861

8.2 CONCLUDING REMARKS

Analyzing the optimal performance achievable in multiobject tracking presents two difficulties: the estimation problems are inherently non-Gaussian and the dimensions are very large. In Sections 3 through 6 we addressed the non-Gaussian problem and in Section 7 we addressed the dimensionality issue. Of the different methods we investigated, three appear promising enough to warrant further research: the Cramer-Rao method of Section 4, the representation theorem method of Section 6, and the random point process approach of Section 7. The Cramer-Rao method of Section 4 has the attractive computational property that its complexity only decreases linearly with time. Furthermore, numerical examples suggest the lower bound is an accurate approximation of steady-state performance in some cases. We need to understand now precisely what these cases are, and prove the validity of the Cramer-Rao bound as a steady-state approximation.

The Monte-Carlo approach of Section 6 is promising but requires further work before we can decide conclusively whether or not it is practically useful. The advantage of this approach is that it can give an accurate estimate of optimal performance, whether in acquisition or steady-state phase -- if one computes long enough. The disadvantage is that this may require impractically long computation. The computational aspects of the problem require further investigation and we indicated some promising directions in Section 6.

In Section 7 we presented a new theoretical approach to analyzing the type of hybrid state problem arising in multiobject tracking. We also sketched how this approach might be used with Cramer-Rao methods to obtain performance bounds. We believe the random point process formulation and techniques of Section 7 will prove useful in analyzing multiobject tracking problems by other methods as well. This approach seems especially fruitful for seeking analytical results as the formulation allows one to represent complex multi-object tracking problems very compactly.

REFERENCES

1. Delaney, J.R., A.S. Willsky, A.L. Blitz, and J. Korn, "Tracking Theory for Airborne Surveillance Radars," ALPHATECH Technical Report TR-142, Burlington, Massachusetts, February 1983.

2. Washburn, R.B., T. Kurien, A.L. Blitz, and A.S. Willsky, "Hybrid State Estimation Approach to Multiobject Tracking for Airborne Surveillance Radars," ALPHATECH Technical Report TR-180, Burlington, Massachusetts, October 1984.

3. Anderson, B.D.O. and J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

4. Van Trees, H.L., Detection, Estimation, and Modulation Theory, Part I, Wiley, New York, 1968.

5. Snyder, D.L. and I.B. Rhodes, "Filtering and Control Performance Bounds with Implications on Asymptotic Separation," Automatica, Vol. 8, pp. 747-753, 1972.

6. Bobrovsky, B.Z. and M. Zakai, "A Lower Bound on the Estimation Error for Certain Diffusion Processes," IEEE Transacations on Information Theory, pp. 45-52, Vol. IT-22, January, 1976.

7. Galdos, J.I., "A Cramer-Rao Bound for Multidimensional Discrete-Time Dynamical Systems," IEEE Transactions on Automatic Control, Vol. AC-25, pp. 117-119, February 1980.

8. Taylor, J.H., "The Cramer-Rao Estimation Error Lower Bound Computation for Deterministic Nonlinear Systems," IEEE Transactions on Automatic Control, Vol. AC-24, pp. 343-344, April 1979.

9. Hammersley, J.M. and D.C. Handscomb, Monte-Carlo Methods, Chapman and Hall, London, 1964.

10. Davis, P.J. and P. Rabinowitz, Methods of Numerical Integration, Academic Press, Orlando, Florida, 1984.

11. Washburn, R.B., "Unified Theory for Airborne Acoustic ASW Processing and Tracking," ALPHATECH Technical Report TR-161, Burlington, Massachusetts, 1983.

12. Washburn, R.B., D. Teneketzis, and A.S. Willsky, "Performance Analysis of Signal Processing and Tracking Systems for Airborne Acoustic ASW," ALPHATECH Technical Report TR-206, Burlington, Massachusetts, July 1984.

13. Vaca, M.V. and D.L. Snyder, "Estimation and Decision for Observations Derived from Martingales: Part II," IEEE Transactions on Information Theory, Vol. IT-24, January 1978.

14. Lindgren, A.G., S.C. Nardone, and K.F. Gong, "Estimating Trajectories in a Dense Environment," Proceedings of the Asilomar Conference on Circuits, Systems, and Computers, 1982.

15. Shannon, C.E., "A Mathematical Theory of Communication," Bell Systems Technical Journal, Vol. 27, pp. 379-423, 623-656, 1948.

16. Berger, T., Rate Distortion Theory, Prentice-Hall, Englewood Cliffs, NJ, 1971.

17. Zakai, M. and J. Ziv, "Lower and Upper Bounds on the Optimal Filtering Error of Certain Diffusion Processes," IEEE Transactions on Information Theory, Vol. IT-18, pp. 325-331, May 1972.

18. Galdos, J.I., "A Lower Bound on Filtering Error with Application to Phase Demodulation," IEEE Transactions on Information Theory, Vol. IT-25, pp. 452-462, July 1979.

19. Galdos, J.I., "A Rate Distortion Theory Lower Bound on Desired Function Filtering Error," IEEE Transactions on Information Theory, Vol. IT-27, pp. 366-368, May 1981.

20. Washburn, R.B. and D. Teneketzis, "Rate Distortion Bound for the Cubic Sensor Problem," ALPHATECH Technical Memorandum TM-162, Burlington, Massachusetts, October 1984.

21. Washburn, R.B. and D. Teneketzis, "Investigation in Performance Analysis on Nonlinear Filtering Problems," ALPHATECH Technical Report TR-264, Burlington, Massachusetts, October 1985.

22. Snyder, D.L. and P.M. Fishman, "How to Track a Swarm of Fireflies by Observing Their Flashes," IEEE Transactions on Information Theory, Vol. IT-21, pp. 692-695, November 1975.

23. Fishman, P.M. and D.L. Synder, "The Statistical Analysis of Space-Time Point Processes," IEEE Transactions on Information Theory, Vol. IT-22, pp. 257-274, May 1976.

24. Halmos, P.R., Measure Theory, Springer-Verlag, New York, 1974.

25. Rudin, W., Real and Complex Analysis, McGraw-Hill, New York, 1966.

26. Kallenberg, D., Random Measures, Academic Press, London, 1976.

27. Wong, E., Stochastic Processes in Information and Dynamical Systems, McGraw-Hill, New York, 1971.

28. Washburn, R.B. and D. Tenetetzis, "Investigations in Performance Analysis of Nonlinear Filtering Problems," ALPHATECH Technical Report TR-264, Burlington, Massachusetts, 1985.

29. Verghese, G. and T. Kailath, "A Further Note on Backwards Markovian Models," IEEE Transactions on Information Theory, Volume IT-25, p. 501, January 1979.

30. LeGrand, F., "Monte-Carlo Methods in Nonlinear Filtering and Importance Sampling," Proceedings of the 23rd IEEE Conference on Decision and Control, December 1984.

APPENDIX A

SECOND-ORDER EXPANSION OF MINIMUM MEAN SQUARE ERROR

As shown in subsection 3.2.2 the minimum mean square error V is given by

$$V = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} - \varepsilon \frac{\sigma^4[\sigma_1^2 - \sigma_0^2]^2}{[\sigma^2 + \sigma_0^2]^2 [\sigma^2 + \sigma_1^2]^2} J \qquad (A-1)$$

where

$$J = \int_{-\infty}^{\infty} \frac{\xi^2 \phi_1(\xi)^2}{(1-\varepsilon)\phi_0(\xi) + \varepsilon \phi_1(\xi)} d\xi \quad , \qquad (A-2)$$

$$\phi_i(\xi) = (2\pi Q_i)^{-1/2} \exp\{-1/2 \xi^2 Q_i^{-1}\} \quad , \qquad (A-3)$$

$$Q_i = \sigma^2 + \sigma_i^2 \quad . \qquad (A-4)$$

We wish to approximate J in the case that $\sigma \approx \sigma_1 \gg \sigma_0$ and ε is small.

Reduction of Integral

By making the substitution

$$\zeta^2 = \frac{\xi^2}{2Q_1}$$

we find that

$$J = \alpha I \qquad (A-5)$$

where I is the integral

$$I = \int_{-\infty}^{\infty} \frac{\zeta^2 \exp\{-\zeta^2\}}{1 + \rho \exp\{-\lambda\zeta^2\}} \, d\zeta$$

$$\alpha = \frac{4 \, Q_1 \, \varepsilon^{-1}}{\sqrt{\pi}} \tag{A-6}$$

$$\rho = \left(\frac{1 - \varepsilon}{\varepsilon} \right) \left(\frac{Q_1}{Q_0} \right)^{1/2} \tag{A-7}$$

$$\lambda = \left(\frac{Q_1}{Q_0} - 1 \right) \quad . \tag{A-8}$$

Using the substitution

$$u = \lambda\zeta^2 - \ell n\rho$$

gives us

$$I = \frac{1}{2} \, \gamma^{3/2} \, |\ell n\rho|^{1/2} \, \rho^{-\gamma} \cdot K \tag{A-9}$$

where the integral K is

$$K = \int_{-\ell n\rho}^{\infty} \frac{\left(1 + \frac{u}{\ell n\rho} \right)^{1/2} e^{-u\gamma}}{1 + e^{-u}} \, du$$

and

$$\gamma = \lambda^{-1} \quad . \tag{A-10}$$

The integral K can be written as the sum

$$K = K_1 + K_2 \tag{A-11}$$

$$K_1 = \int_0^{\infty} \frac{\left(1 + \frac{u}{\ell n \rho}\right)^{1/2} e^{-u\gamma}}{1 + e^{-u}} \, du$$

and K_2 is the integral

$$K_2 = \int_{-\ell n \rho}^{0} \frac{\left(1 + \frac{u}{\ell n \rho}\right)^{1/2} e^{-u\gamma}}{1 + e^{-u}} \, du$$

The second integral K_2 can also be expressed as

$$K_2 = \ell n \rho \cdot \int_0^1 \frac{(1-z)^{1/2} \cdot e^{z \cdot \theta}}{1 + e^{-2\ell n \rho}} \, dz$$

where

$$\theta = (1-\gamma)\ell n \rho \quad . \tag{A-12}$$

In the case of interest to us, $\gamma \approx 1$, $\ell n \rho \gg 1$, but θ is of order 1. We will now approximate K_1 and K_2 under these conditions.

Approximation of K_1 and K_2

Write

$$a = \ell n \rho \quad , \tag{A-13}$$

so that $K_1$and K_2 become

$$K_1 = \int_{\infty}^{0} \frac{\left(1 + \frac{u}{a}\right)^{1/2} e^{-\gamma u}}{1 + e^{-u}} \, du$$

$$K_2 = a \cdot \int_{\infty}^{0} \frac{(1-z)^{1/2} \, e^{\theta z}}{1 + e^{-z}} \, dz$$

where we assume $a > 0$, $\gamma \approx 1$, and $\theta \sim 1$. We approximate K_1 as

$$K_1 = K_{11} + K_{12}$$

where

$$K_{11} = \beta(\gamma) = \int_0^\infty \frac{e^{-\gamma u}}{1 + e^{-u}}\, du$$

is the incomplete beta function and

$$K_{12} = \int_0^\infty \frac{\left(1 + \frac{u}{a}\right)^{1/2} - 1}{1 + e^{-u}}\, du \quad .$$

The second term K_{12} can be bounded by

$$0 < K_{12} < \frac{1}{2a} \int_0^\infty u e^{-\gamma u}\, du = \frac{1}{2a\ \gamma^2} \quad .$$

Note that

$$\beta(1) = \ell n 2 \quad ,$$

$$\beta'(1) = \frac{-\pi^2}{12} \quad ,$$

and

$$\beta''(1) < \int_0^\infty u^2 e^{-\gamma u}\, du = \frac{2}{\gamma^3} \quad .$$

Thus, we have

$$\beta(\gamma) = \ell n 2 - \frac{\pi^2}{12}(\gamma - 1) + \frac{1}{2}\Delta(\gamma) \cdot (1-\gamma)^2$$

where

$$\Delta(\gamma) \leq \frac{2}{\gamma^3} \quad \text{if } \gamma < 1$$

$$\Delta(\gamma) \leq 2 \quad \text{if } \gamma \geq 1 \quad .$$

Thus, we approximate K_1 by

$$K_1 \approx \ell n 2 - \frac{\pi^2}{12}(\gamma-1) \tag{A-14}$$

where the magnitude of the error is bounded by

$$\frac{1}{2a\ \gamma^2} + (1-\gamma)^2 \max\{\gamma^{-3},\ 1\} \quad . \tag{A-15}$$

For K_2 we use the crude bounds

$$\frac{1}{3} \leq a^{-1}K_2 \leq \frac{2}{3}\ e^{\theta} \tag{A-16}$$

if $\theta \geq 0$, or

$$\frac{1}{3}\ e^{\theta} \leq a^{-1}K_2 \leq \frac{2}{3} \quad . \tag{A-17}$$

if $\theta \leq 0$.

Upper and Lower Bounds

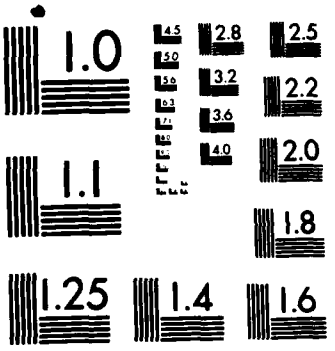
Let us assume $\gamma > 1$; this is equivalent to assuming

$$\sigma_1^2 - \sigma^2 < 2\sigma_0^2 \quad . \tag{A-18}$$

Furthermore, assume that $\ell n\rho > 0$. This is equivalent to

AD-A162 225 PERFORMANCE ANALYSIS FOR HYBRID STATE ESTIMATION PROBLEMS(U) ALPHATECH INC BURLINGTON MA R B WASHBURN ET AL OCT 85 N00014-84-C-0370 2/2

UNCLASSIFIED F/G 12/1 NL



$$\varepsilon < \left(1 + e \left[\frac{Q_0}{Q_1} \right]^{1/2} \right)^{-1} \quad . \tag{A-19}$$

Note that θ given by Eq. A-12 is negative. In this case, K is bounded above and below by

$$K^- < K < K^+ \tag{A-20}$$

where

$$K^- = \ell n 2 - \frac{\pi^2}{12} (\gamma - 1) - \frac{1}{2\gamma^2 \ell n \rho} - (1-\gamma)^2 + \frac{1}{3} e^{\theta} a \tag{A-21}$$

$$K^+ = \ell n 2 - \frac{\pi^2}{12} (\gamma - 1) + \frac{1}{2\gamma^2 \ell n \rho} + (1-\gamma)^2 + \frac{2}{3} \cdot a \quad . \tag{A-22}$$

The corresponding bounds on J are then

$$J^- < J < J^+ \tag{A-23}$$

where

$$J^{\pm} = \alpha \gamma^{3/2} |\ell n \rho|^{1/2} \rho^{-\gamma} K^{\pm} \quad .$$

Thus, one gets the upper and lower bounds $V_2^{\pm}$

$$V_2^{\pm} = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2} + \varepsilon \frac{\sigma^4 [\sigma_1^2 - \sigma_0^2]}{[\sigma^2 + \sigma_0^2]^2} - \varepsilon^2 \frac{\sigma^4 [\sigma_1^2 - \sigma_0^2]^2}{[\sigma^2 + \sigma_0^2]^2 [\sigma^2 + \sigma_1^2]^2} J^{\mp} \tag{A-25}$$

Asymptotic Behavior as $\varepsilon \to 0$

Assume that $\gamma > 1$ is maintained as before and let $\varepsilon \to 0$. Then note that

$$K_2 = \int_0^1 \frac{(1-z)^{1/2} \ell n \rho \, e^{-z \delta \ell n \rho} \, dz}{1 + e^{-z \ell n \ell}}$$

where

$$\delta = \gamma - 1 > 0 \quad . \tag{A-26}$$

Thus,

$$\lim_{\varepsilon \to 0} K_2 = 0 \quad .$$

Furthermore,

$$\lim_{\varepsilon \to 0} K_1 = \beta(\gamma) \quad .$$

Note that as $\varepsilon \to 0$

$$\rho \sim \varepsilon^{-1} \left(\frac{Q_1}{Q_0} \right)^{1/2} \quad ,$$

$$\ell n \rho \sim \ell n \ \varepsilon \quad ,$$

$$\alpha \sim \varepsilon^{-1} \cdot \frac{4 \ Q_1}{\sqrt{\pi}} \quad .$$

It follows that as $\varepsilon \to 0$

$$J \sim c \cdot \varepsilon^{\gamma - 1} \ | \ell n \ \varepsilon |^{1/2} \tag{A-27}$$

where

$$c = \frac{4 \ Q_1}{\sqrt{\pi}} \gamma^{3/2} \left(\frac{Q_1}{Q_0} \right)^{- \ 1/2\gamma} \beta(\gamma) \quad . \tag{A-28}$$

In this case, J does approach 0 as $\varepsilon \to 0$ but not very rapidly since $\gamma - 1$ is assumed small.

END

FILMED

1-86

DTIC